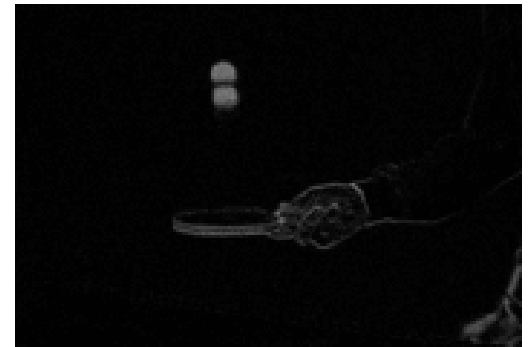
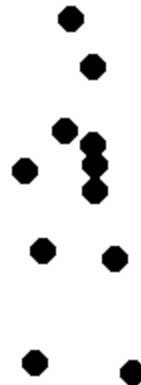
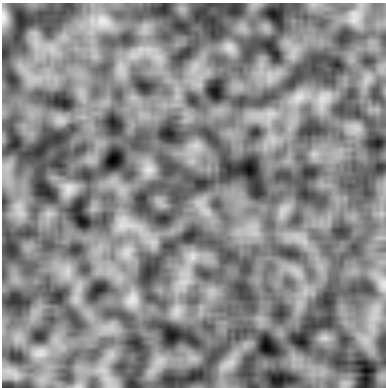


Analisis del movimiento

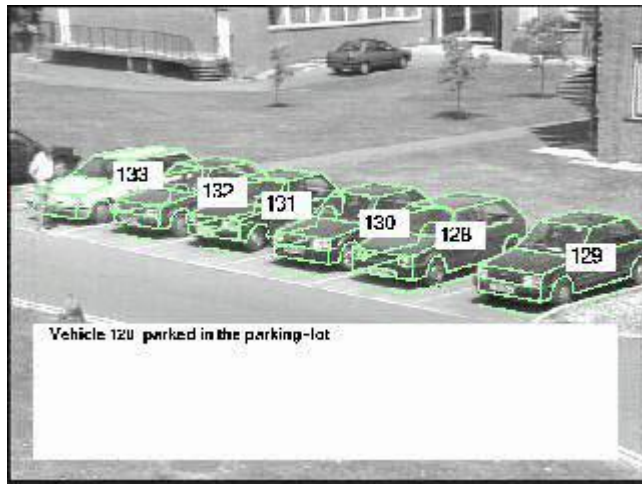
Estimación & Seguimiento

Antoine MANZANERA – ENSTA/LEI



Aplicaciones (1)

Vigilancia de aldea



(University of Surrey)

Compresión e Indexación



Aplicaciones (2)

Navegación robótica

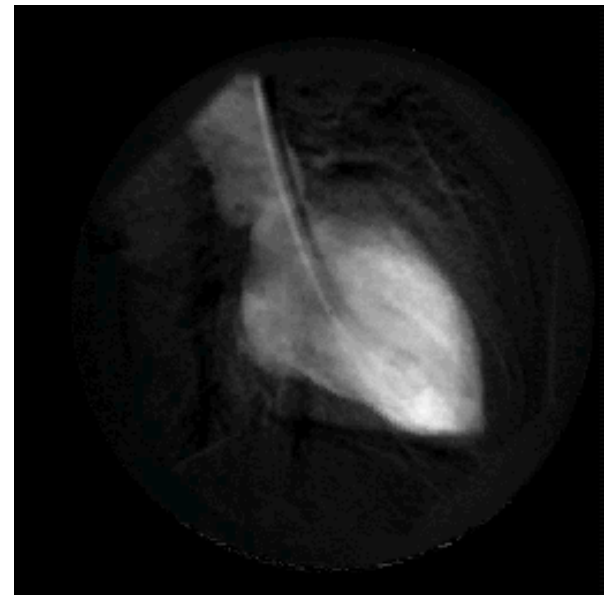


(Ecole des Mines / CMM)

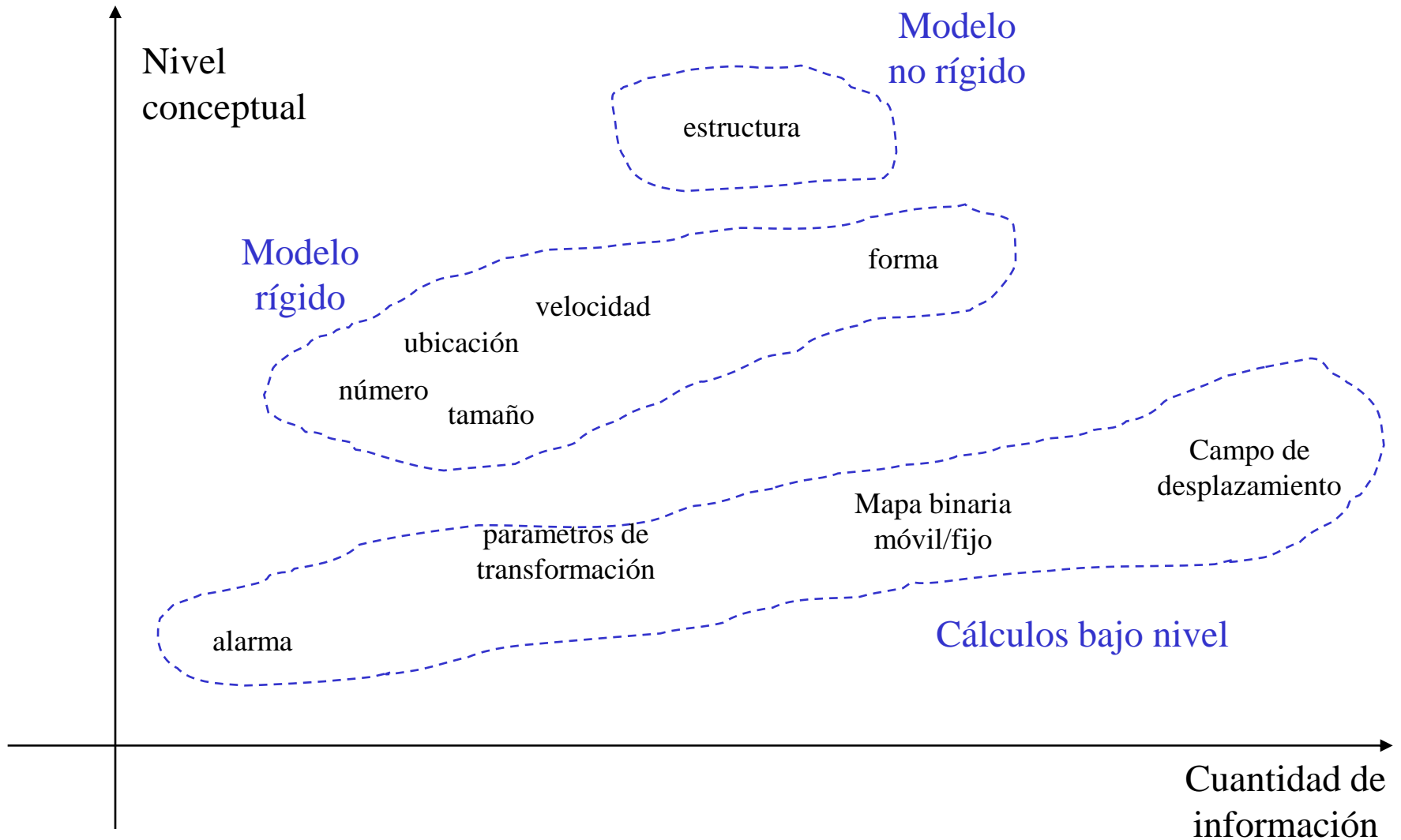
Seguimiento automático



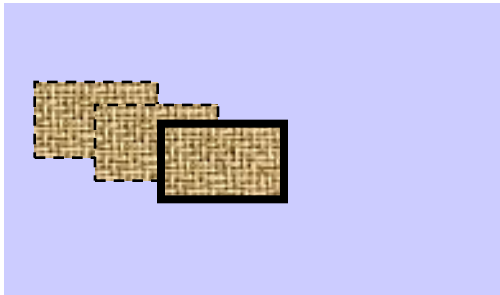
Movimiento fluido



¿Qué información extraer?



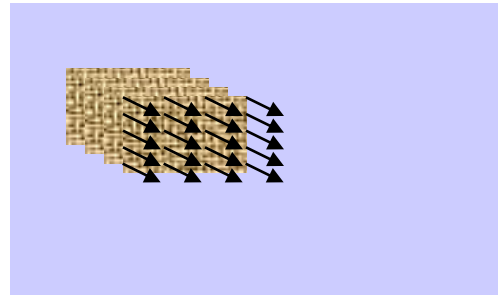
DetECCIÓN – Estimación – Persecución



DETECCIÓN

Objetivo: Identificar en cada imagen, píxeles que pertenecen a objetos móviles

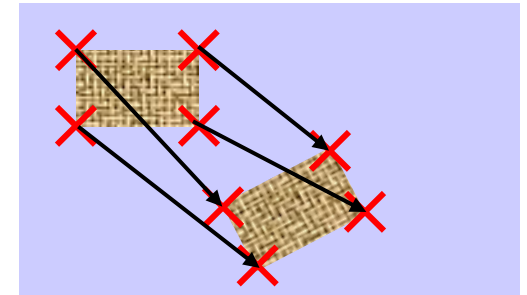
- Cierta continuidad temporal
- Movimiento de la cámara cero o muy sencillo.



ESTIMACIÓN

Objetivo: calcular el movimiento aparente (velocidad instantánea) de cada píxel

- Continuidad temporal
- « procesar después »



seguimiento

Objetivo: aparear ciertas estructuras espaciales por cada par de imágenes.

- Discontinuidad temporal
- « procesar antes »



Capítulo 1



Capítulo 2

ÍNDICE CAPÍTULO

- **Campo de velocidades aparentes y aplicaciones**
- **Técnicas por apareamiento**
- **Técnicas diferenciales 1: Lucas y Kanade**
- **Técnicas diferenciales 2: Horn y Schunck**

Cálculo del movimiento aparente

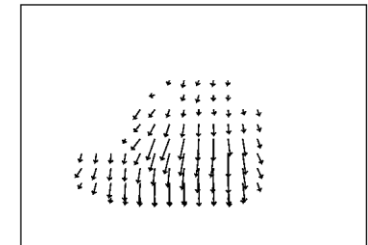
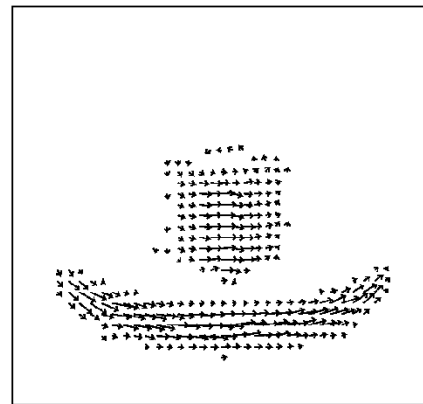
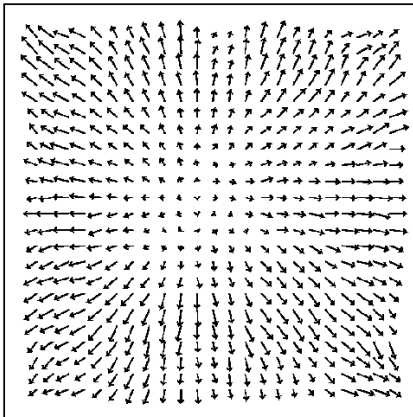
- (1) El cálculo de un movimiento aparente *global* (mapeo) entre dos imágenes corresponde a la estimación de parámetros de una transformación que afecta a todos los puntos de la imagen: traslación, rotación, escalamiento, afinidad, ...
- (2) El cálculo de un movimiento aparente *local* consiste en asociar a cada píxel (x,y,t) de I un vector (v_x^t, v_y^t) que representa la *velocidad aparente* del píxel (x,y) en el instante t .

—————→ Cálculo del *flujo óptico* (= Campo de movimiento)

Idealmente: el vector (v_x^t, v_y^t) representa la proyección sobre el plano imagen del vector velocidad (V_x^t, V_y^t, V_z^t) de los objetos de la escena respecto a un sistema de referencia imagen (O, x, y, z) .

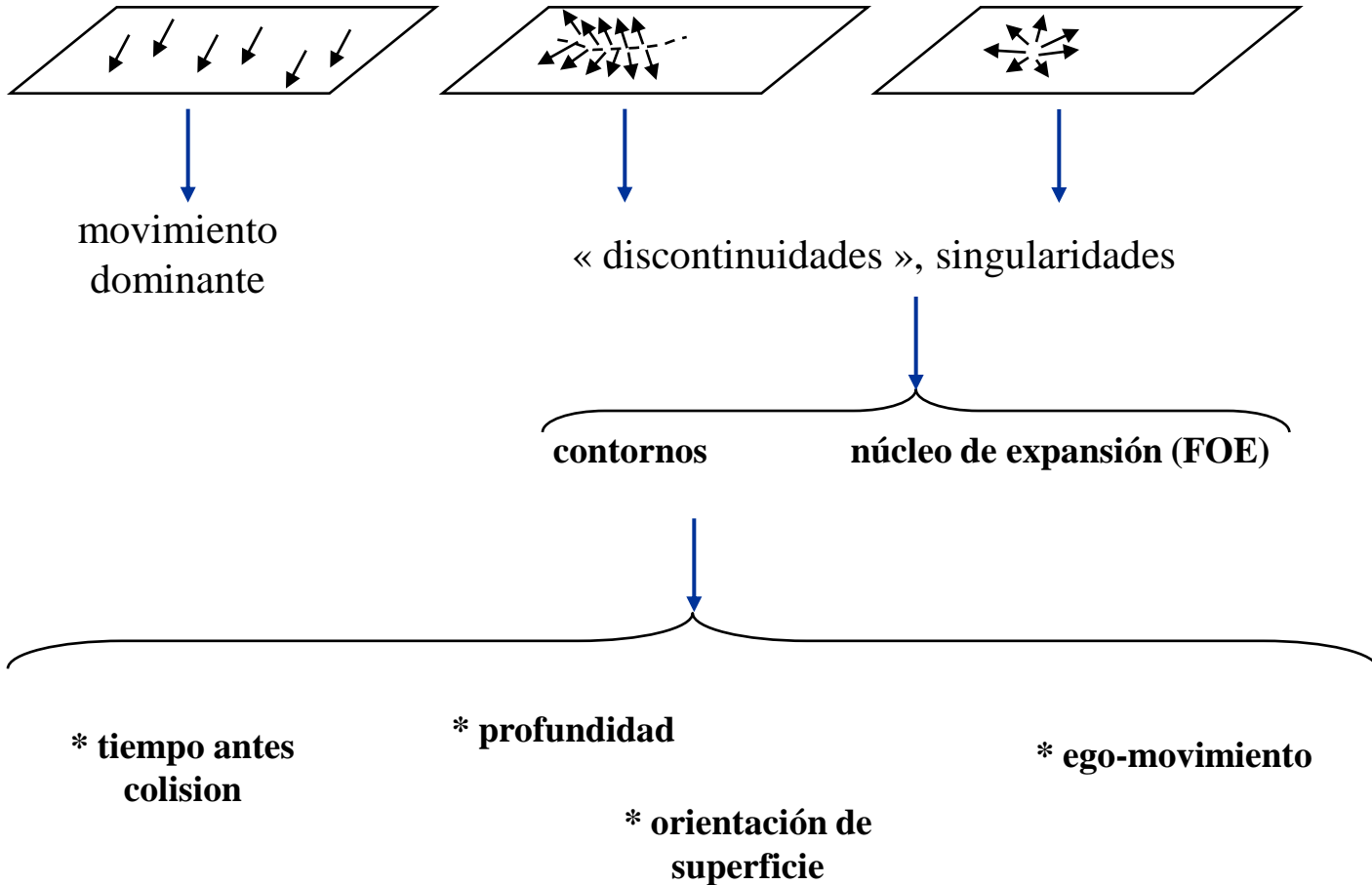
Será calculado a partir de las variaciones temporales de la función $I(x,y,t)$.

Ejemplos de flujo óptico



Fuente: Pierre Kornprobst - INRIA

¿ Cuales informaciones puede proveer el flujo óptico ?



Nivel conceptual

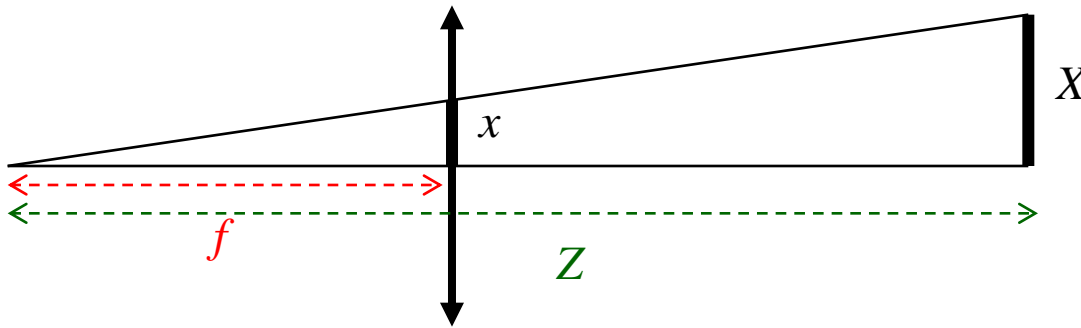
Distorsión perspectiva

Se denota (X,Y,Z) las coordenadas de un punto de la escena.

(x,y) las coordenadas de la proyección del mismo en la imagen.

f la distancia focal de la cámara

***Distorsión perspectiva
(modelo estenopeico) :***



$$x = f \frac{X}{Z}$$
$$y = f \frac{Y}{Z}$$

Movimiento proyectivo

Supongamos que la cámara se mueve a la velocidad $(-X', -Y', -Z')$ respecto a una escena estática (escena no deformable, sin objetos móviles), entonces todos los puntos de la escena tienen misma velocidad (X', Y', Z') respecto a la cámara (con $X' = dX/dt$; $Y' = dY/dt$; $Z' = dZ/dt$). Derivando respecto al tiempo las ecuaciones de distorsión perspectiva, obtenemos:

$$\begin{cases} x' = f \left(\frac{X'}{Z} - \frac{XZ'}{Z^2} \right) \\ y' = f \left(\frac{Y'}{Z} - \frac{YZ'}{Z^2} \right) \end{cases} \quad \text{Soit : } \begin{pmatrix} x' \\ y' \end{pmatrix} = \frac{1}{Z} \begin{pmatrix} -f & 0 & x \\ 0 & -f & y \end{pmatrix} \begin{pmatrix} -X' \\ -Y' \\ -Z' \end{pmatrix}$$

Un modelo más realista...

En realidad, el centro óptico (proyección del diafragma sobre el plano imagen) no está generalmente en el centro del imagen sino en un punto (x_c, y_c) . Además, la distorsión perspectiva no es generalmente la misma en ambas direcciones, y la distancia focal se representa con el par (f_x, f_y) :

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \frac{1}{Z} \begin{pmatrix} -f_x & 0 & (x - x_c) \\ 0 & -f_y & (y - y_c) \end{pmatrix} \begin{pmatrix} -X' \\ -Y' \\ -Z' \end{pmatrix}$$

Los parametros $\{ x_c, y_c, f_x, f_y \}$ dependen del sensor y de la óptica. Son estimados en una fase de *calibración* de la cámara.

Por otro lado, si se toma en cuenta el componente de rotación $\Omega = (\Omega_x, \Omega_y, \Omega_z)$ de la cámara respecto a la escena:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \frac{1}{Z} \begin{pmatrix} -f_x & 0 \\ 0 & -f_y \end{pmatrix} \begin{pmatrix} x - x_c \\ y - y_c \end{pmatrix} + \begin{pmatrix} \frac{(x-x_c)(y-y_c)}{f_y} & f_y + \frac{(y-y_c)^2}{f_y} & \frac{f_x}{f_y}(y-y_c) \\ f_y + \frac{(y-y_c)^2}{f_y} & -\frac{(x-x_c)(y-y_c)}{f_x} & -(x-x_c) \end{pmatrix} \begin{pmatrix} \Omega_x \\ \Omega_y \\ \Omega_z \end{pmatrix}$$

factores de perspectiva
coordenadas del centro óptico
componente de traslación
componente de rotación

Traslación pura y profundidad

- Traslación pura de la cámara según el eje x , sea $T = (-X', 0, 0)$ y $\Omega = (0, 0, 0)$

$$\begin{cases} x' = f_x \frac{X'}{Z} \\ y' = 0 \end{cases} \quad \text{sea} \quad \begin{cases} Z = f_x \frac{X'}{x'} \end{cases}$$

→ Velocidades aparentes horizontales, de módulos inversamente proporcionales a la profundidad.

- Traslación pura de la cámara según el eje z , sea $T = (0, 0, -Z')$ y $\Omega = (0, 0, 0)$

$$\begin{cases} x' = \frac{(x - x_c)Z'}{Z} \\ y' = \frac{(y - y_c)Z'}{Z} \end{cases} \quad \text{sea} \quad \begin{cases} Z = \frac{(x - x_c)Z'}{x'} \\ Z = \frac{(y - y_c)Z'}{y'} \end{cases}$$

→ Zoom sobre la imagen con núcleo de expansión en el centro óptico.

El núcleo de expansión (FOE)

Durante la traslación de la cámara en una escena estática, las direcciones de velocidad de los puntos proyectados sobre el plano imagen, convergen hacia un punto del plano proyectivo denominado FOE.

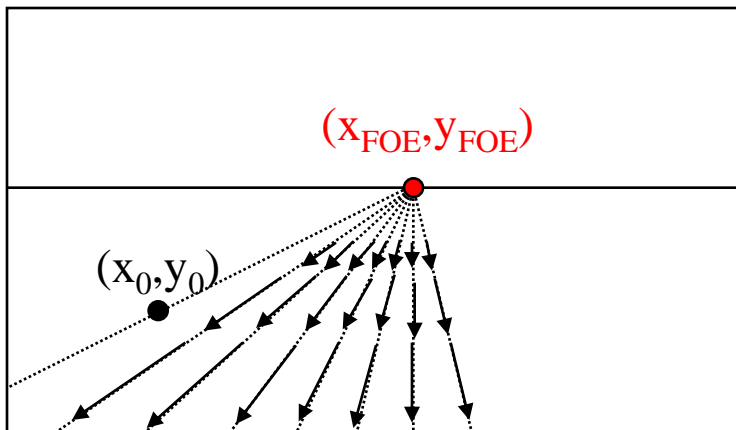
En lo siguiente, se supone que la cámara se mueve según la traslación $T = (-X', -Y', -Z')$. Para simplificar las notaciones, se supone además que $f_x = f_y = f$ et $(x_c, y_c) = (0, 0)$.

Sea (X_0, Y_0, Z_0) un punto de la escena. Después de un tiempo t , esta proyectado sobre la imagen en el punto (x_t, y_t) , con:

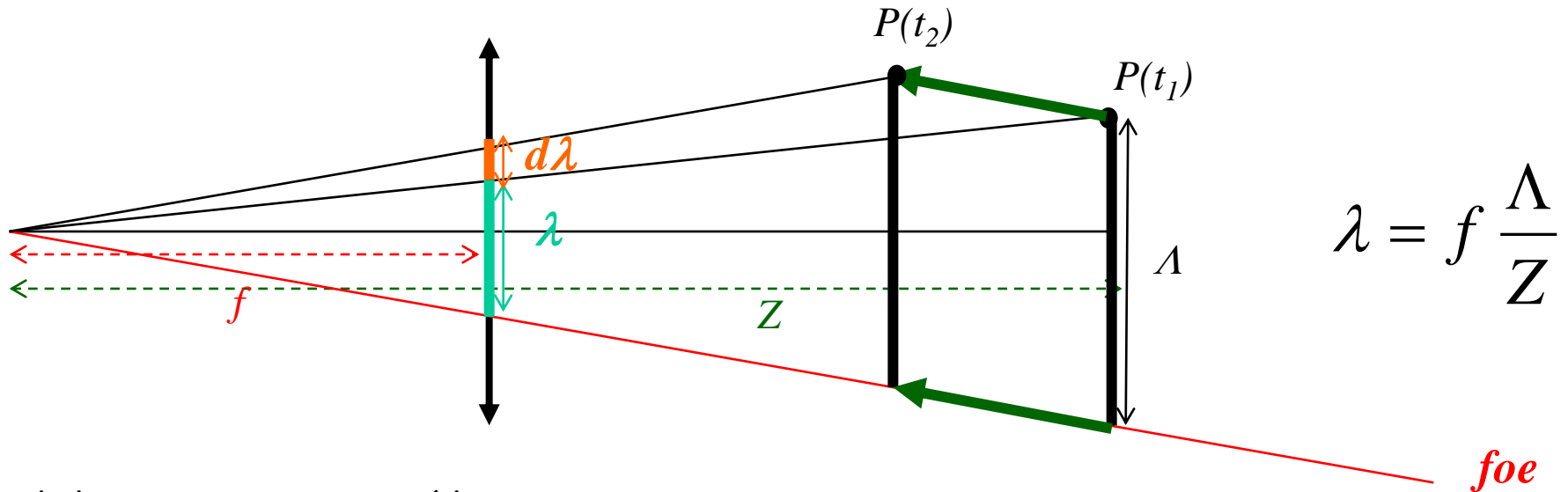
$$(x_t, y_t) = \left(f \frac{X_0 + tX'}{Z_0 + tZ'}, f \frac{Y_0 + tY'}{Z_0 + tZ'} \right)$$

$$(x_{FOE}, y_{FOE}) = \lim_{t \rightarrow -\infty} (x_t, y_t)$$

$$= \left(f \frac{X'}{Z'}, f \frac{Y'}{Z'} \right)$$



Tiempo antes colisión



movimiento en una escena estática:

$$\lambda' = -\frac{f\Lambda}{Z^2} Z'$$

y pues

$$\frac{\lambda}{\lambda'} = -\frac{Z}{Z'}$$

tiempo
antes
colisión

Cálculo del flujo óptico: limitaciones

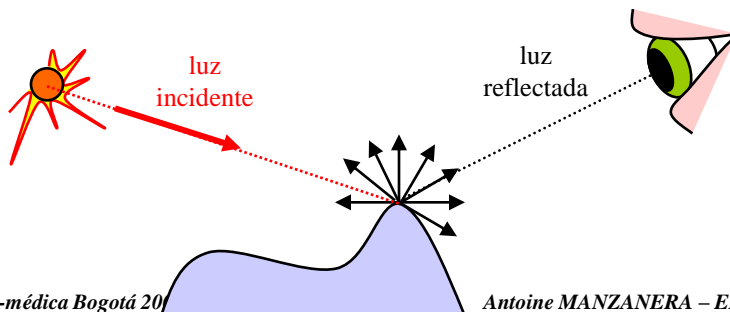


(1) Se supone:

MOVIMIENTO \Leftrightarrow VARIACIÓN DE INTENSIDAD

(\Leftarrow) Se supone que una variación de intensidad observada en la imagen traduce la presencia de un movimiento en la escena. Esto corresponde a una hipótesis de *iluminación constante* en la escena.

(\Rightarrow) Es el cambio de intensidad en la imagen que permite recoger el movimiento aparente del objeto. Se supone entonces que la luz reflejada por cada punto del objeto se mantiene constante independientemente del movimiento relativo objeto/cámara. La hipótesis subyacente es que los objetos son de *superficie lambertiana*:



La intensidad de la luz reflectada es constante en todas las direcciones.

Cálculo del flujo óptico: limitaciones

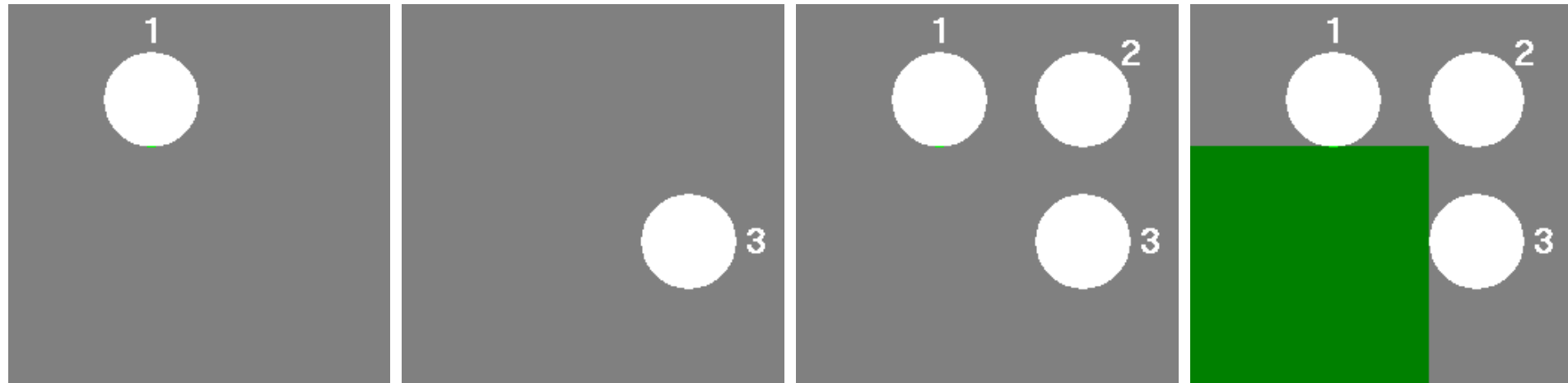


(2) PROBLEMA DE LA APERTURA

El movimiento aparente de un punto se recoge gracias a un cálculo realizado en una vecindad limitada de este punto.



Solo se puede calcular el componente de movimiento *en la dirección del gradiente* (es decir, *perpendicular al contorno*).

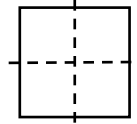


Fuente : PROJET TELESUN <http://telesun.insa-lyon.fr>

Técnica por apareamiento

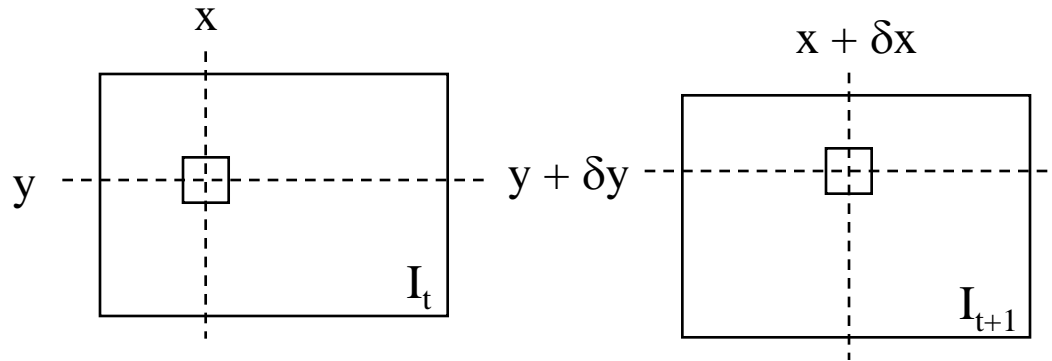
Medida de apareamiento:

Se considera $B \subset \mathbf{Z}^2$ vecindad del origen:



$$A_{(x,y)}^t(\delta x, \delta y) = \sum_{(b_1, b_2) \in B} (I(x + b_1, y + b_2, t) - I(x + \delta x + b_1, y + \delta y + b_2, t + 1))^2$$

Es la suma de diferencias cuadradas (SSD) entre 2 bloques:



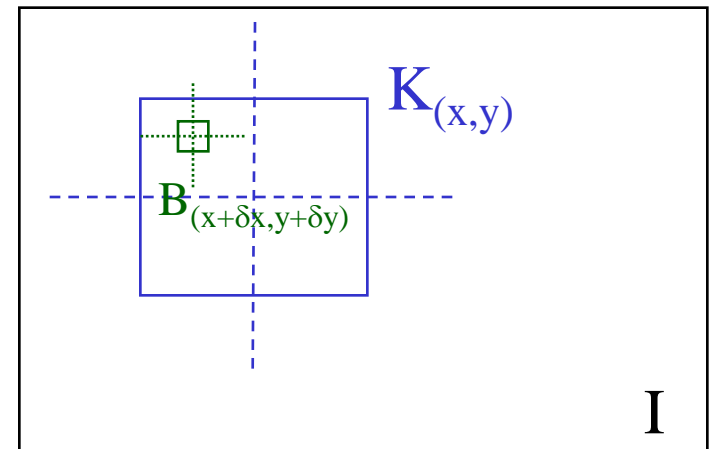
tipicamente: B de tamaño $9 \times 9, 15 \times 15 \dots$

Solución del flujo óptico:

$$(v_x^t, v_y^t) = \arg \min_{(\delta x, \delta y) \in K} A_{(x,y)}^t(\delta x, \delta y)$$

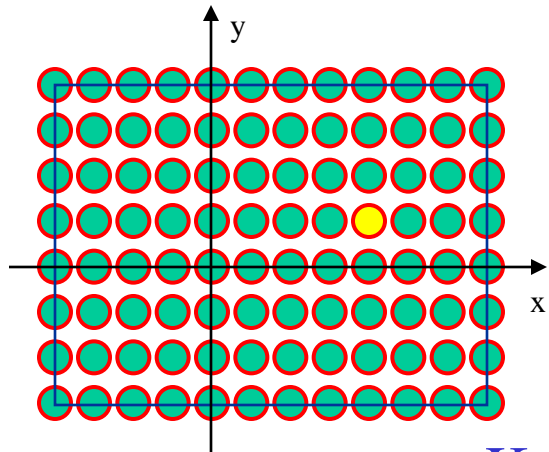
$$K \subsetneq I$$

para limitar el tiempo de calculo
+ Estrategias de optimización...



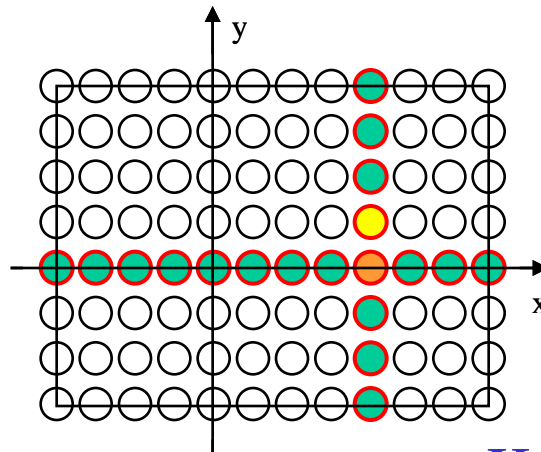
Apareamiento: estrategias de optimización

El principio de las estrategias de optimización es explorar de manera diferente el espacio de búsqueda K para minimizar la complejidad de cálculo del mínimo de la función de apareamiento:



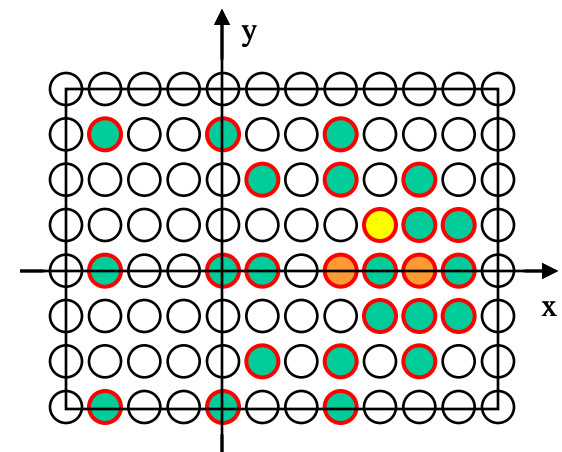
EXHAUSTIVA

K



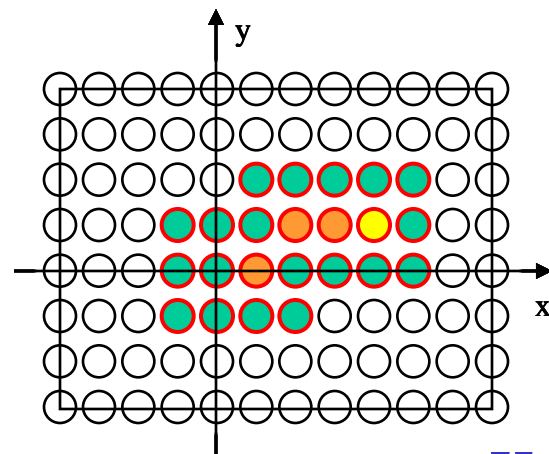
SEPARADA

K



JERÁRQUICA

K



MÉTODO DEL GRADIENTE

K

Para cada punto de K , se calcula la suma de diferencias mientras que se mantiene por debajo de un cierto umbral T . La solución corresponde al punto de K por el cual más diferencias han sido sumadas.

SECUENCIAL

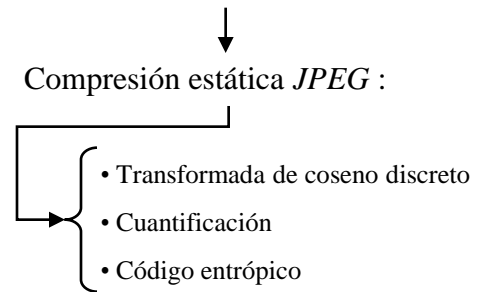
- Estas técnicas no son excluyentes y pueden combinarse.

- Sin hipótesis sobre la función de apareamiento (separabilidad, convexidad ...) en el espacio K , sólo la búsqueda exhaustiva asegura el cálculo del óptimo global sobre K .

Ejemplo: normas mpeg de compresión vídeo



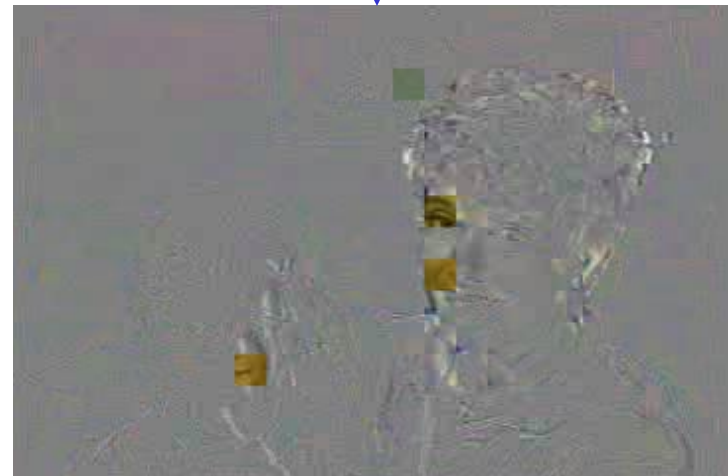
- Cada imagen esta cortada en bloques 16×16 ou 32×32 .
- Hipótesis de mismo desplazamiento para todos los píxeles de cada bloque.
- Codificación: Vectores de desplazamiento + errores comprimidos...



Flujo óptico submuestreado



Trama predicha



Errores de predicción

Fuente :
VcDemo
TU Delft

Técnicas diferenciales (1)

LUCAS & KANADE 1981

El principio del método de Luca y Kanade es de calcular el mínimo de la función de apareamiento cuadrático (SSD), suponiendo que el desplazamiento buscado es pequeño, como el punto donde las derivadas de la función de apareamiento se anulan, respecto a δ_x y a δ_y .

$$\begin{aligned}(v_x^t, v_y^t) &= \arg \min_{(\delta x, \delta y) \in K} A(\delta x, \delta y) \\ &= \arg \min_{(\delta x, \delta y) \in K} \sum_{(x, y) \in B} (I(x, y, t) - I(x + \delta x, y + \delta y, t + 1))^2\end{aligned}$$

Bajo la hipótesis de que I es regular y que el desplazamiento (δ_x, δ_y) es pequeño, se escribe la fórmula de Taylor al orden 1 de I :

$$I(x + \delta x, y + \delta y, t + 1) \approx I(x, y, t) + \frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial t}$$

La función de apareamiento A se vuelve:

$$A(\delta x, \delta y) \approx \sum_{(x, y) \in B} \left(\frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial t} \right)^2$$

Técnicas diferenciales (1)

LUCAS & KANADE 1981

$$A(\delta x, \delta y) \approx \sum_{(x,y) \in B} \left(\frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial t} \right)^2$$

Ecuaciones de Euler-Lagrange de minimización de A: Anulación de las derivadas primeras respecto a δ_x y a δ_y :

$$2 \sum_{(x,y) \in B} \left(\frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial t} \right) \frac{\partial I}{\partial x} = 0$$

$$2 \sum_{(x,y) \in B} \left(\frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial t} \right) \frac{\partial I}{\partial y} = 0$$

Lo que conduce a la resolución del sistema lineal:

$$\mathbf{H} \cdot \mathbf{v} = \mathbf{b}$$

$$\text{con } \mathbf{H} = \begin{pmatrix} \sum_{(x,y) \in B} \left(\frac{\partial I}{\partial x} \right)^2 & \sum_{(x,y) \in B} \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} \\ \sum_{(x,y) \in B} \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} & \sum_{(x,y) \in B} \left(\frac{\partial I}{\partial y} \right)^2 \end{pmatrix} \text{ y } \mathbf{b} = \begin{pmatrix} - \sum_{(x,y) \in B} \frac{\partial I}{\partial x} \frac{\partial I}{\partial t} \\ - \sum_{(x,y) \in B} \frac{\partial I}{\partial y} \frac{\partial I}{\partial t} \end{pmatrix}$$

y $\mathbf{v} = (v_x, v_y)$ el desplazamiento buscado.

La resolución del sistema (S) está finalmente realizada por un método iterativo, tipo Gauss-Seidel:

Inicialización: $v_x^{(0)} = 0; v_y^{(0)} = 0$

$$\text{Para } k > 0: \begin{cases} v_x^{(k)} = \frac{1}{\mathbf{H}_1^1} (\mathbf{b} - \mathbf{H}_1^2 \cdot v_y^{(k-1)}) \\ v_y^{(k)} = \frac{1}{\mathbf{H}_2^2} (\mathbf{b} - \mathbf{H}_2^1 \cdot v_x^{(k)}) \end{cases}$$

Técnicas diferenciales (1)

LUCAS & KANADE 1981

La existencia y la estabilidad de una solución al sistema (S) depende de la matriz \mathbf{H} :

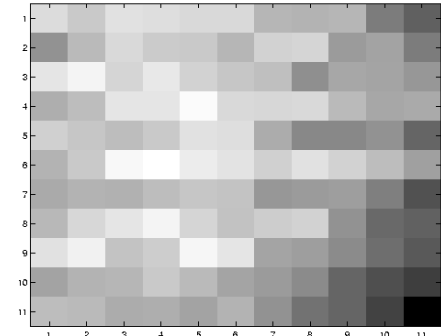
$$\mathbf{H} = \begin{pmatrix} \sum_{(x,y) \in B} \left(\frac{\partial I}{\partial x}\right)^2 & \sum_{(x,y) \in B} \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} \\ \sum_{(x,y) \in B} \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} & \sum_{(x,y) \in B} \left(\frac{\partial I}{\partial y}\right)^2 \end{pmatrix}$$

NB : Encontraremos de nuevo esta matriz en el capítulo 2 para el cálculo de los puntos de interés...

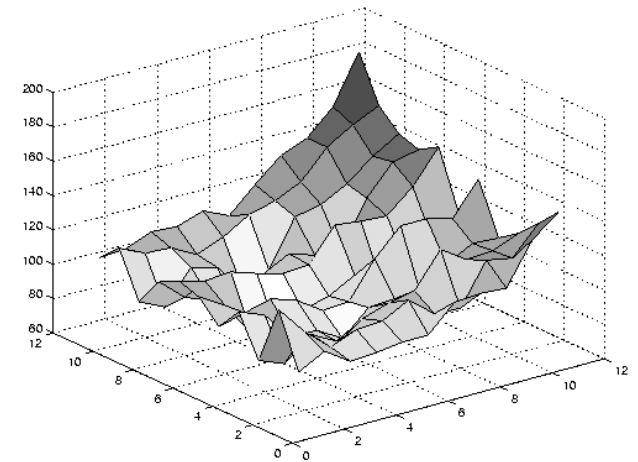
La matriz \mathbf{H} debe ser de rango 2 « en el sentido fuerte », es decir poseer 2 autovalores (λ_1, λ_2) grandes. Los autores proponen usar $\lambda = \min(\lambda_1, \lambda_2)$ como índice de confianza en el desplazamiento calculado.

Esta propiedad corresponde a una interpretación algebraica del *problema de la apertura*...

Rango 0: zona homogénea



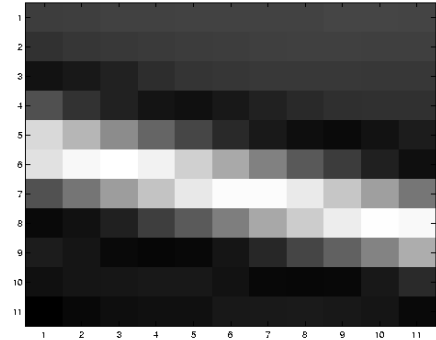
Zona con poca textura:
2 autovalores débiles



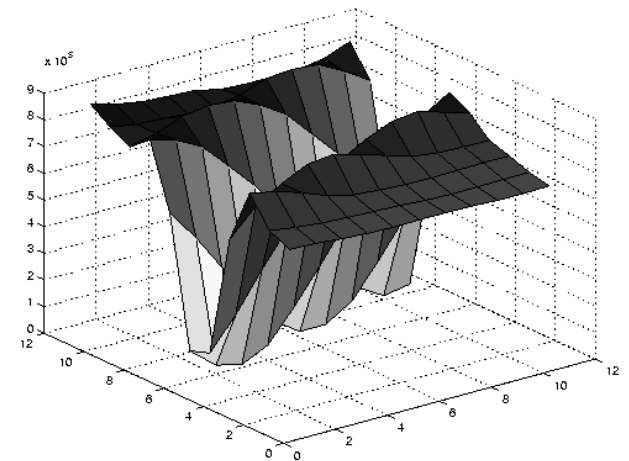
Función de apareamiento A

Fuentes : Steven Seitz UW-Seattle

Rango 1: contorno simple



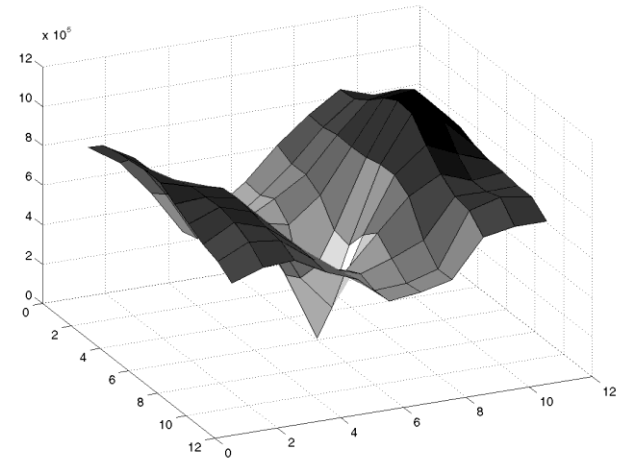
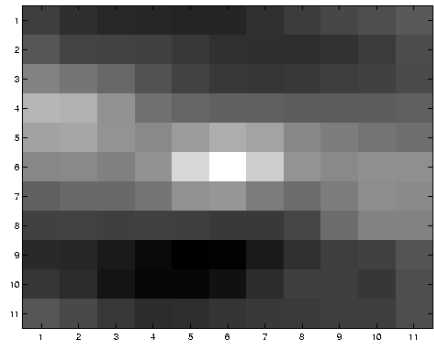
Contorno recto: 1 única
autovalor grande



Fuentes : Steven Seitz UW-Seattle

Función de apareamiento A

Rango 2: punto anguloso



Punto anguloso: 2
autovalores grandes

Fuentes : Steven Seitz UW-Seattle

Función de apareamiento A

Técnicas diferenciales (2)

HORN & SCHUNCK 1981

Explotación directa de la limitación (1) :

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t)$$

Formula de Taylor

$$= I(x, y, t) + \frac{\partial I}{\partial x} \cdot \delta x + \frac{\partial I}{\partial y} \cdot \delta y + \frac{\partial I}{\partial t} \cdot \delta t + \dots$$

Términos de ordenes superiores abandonados

ENTONCES:

$$\nabla I \cdot \mathbf{v} + \frac{\partial I}{\partial t} = 0$$

Ecuación del movimiento aparente (EMA)

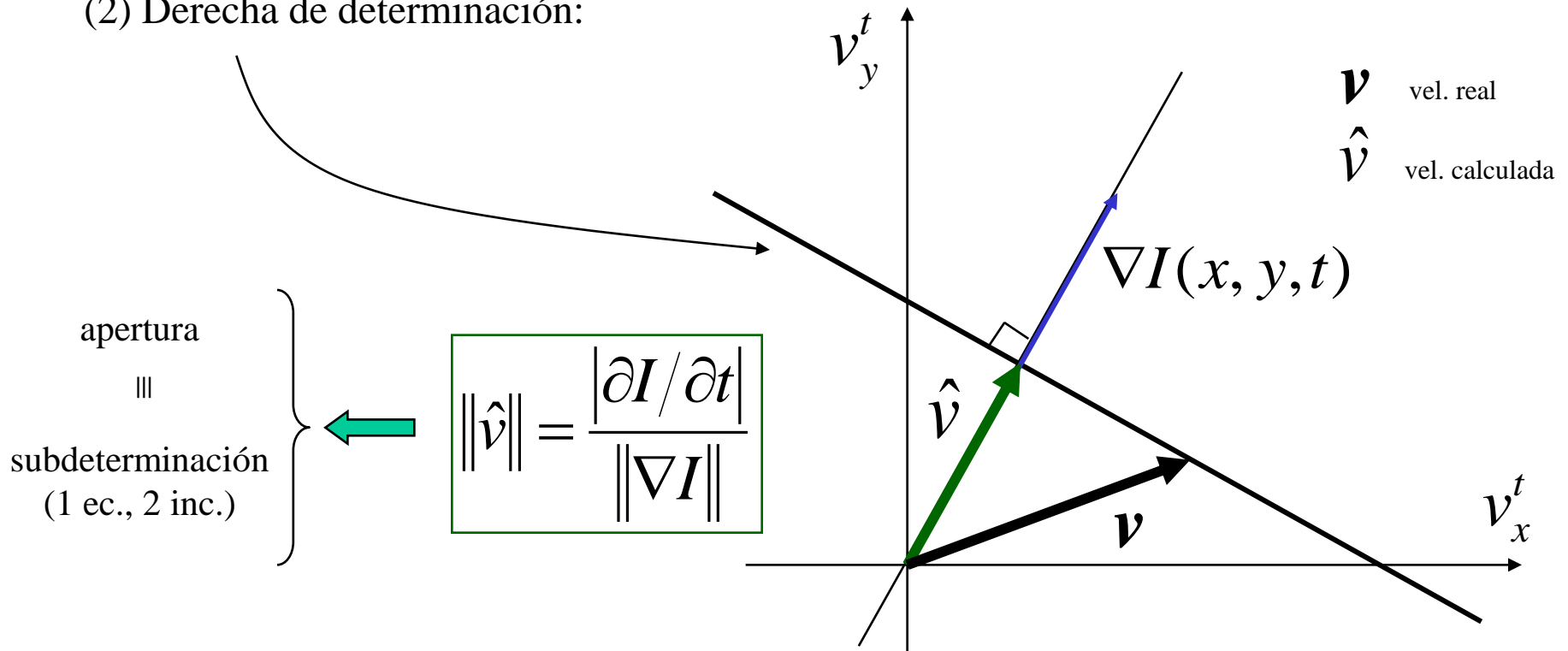
o: ecuación del flujo óptico

$$\text{con } \left\{ \begin{array}{l} \nabla I = \left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right) \text{ gradiente espacial} \\ \mathbf{v} = (v_x^t, v_y^t) \text{ incognitas} \\ \frac{\partial I}{\partial t} \text{ gradiente temporal} \end{array} \right.$$

Interpretación de la EMA

(1) Bajo la hipótesis de cierta *regularidad* del campo y de *pequeños desplazamientos*, los cambios temporales son *equivalentes* (a la primera orden) al producto escalar de los *cambios espaciales* y de la *velocidad aparente*.

(2) Derecha de determinación:



Resolución de la EMA

HORN & SCHUNCK 1981

Resolución de la EMA por *agregación de una hipótesis de regularidad*.

⇔ *Regularización* del problema mal puesto por hipótesis de campo de desplazamiento *suave*.

➔ Minimización de una función de costo:

$$C_{(x,y)}^t(v_x^t, v_y^t) = \underbrace{\left(\nabla I \cdot v + \frac{\partial I}{\partial t} \right)^2}_{\text{EMA}} + \lambda \underbrace{\left[\left(\frac{\partial v_x^t}{\partial x} \right)^2 + \left(\frac{\partial v_x^t}{\partial y} \right)^2 + \left(\frac{\partial v_y^t}{\partial x} \right)^2 + \left(\frac{\partial v_y^t}{\partial y} \right)^2 \right]}_{\text{REGULARIZACIÓN}}$$

EMA

REGULARIZACIÓN

factor de
ponderación

Resolución de la EMA

HORN & SCHUNCK 1981

1. Minimización de una función cuadrática:

$$C_{(x,y)}(u, v) = (I_x u + I_y v + I_t)^2 + \lambda(u_x^2 + u_y^2 + v_x^2 + v_y^2)$$

$$\text{con: } \begin{cases} (u, v) \text{ componentes del campo (desconocidos) a calcular} \\ \{I_x, I_y, I_t\} \text{ derivadas parciales de la imagen} \\ \{u_x, u_y, v_x, v_y\} \text{ derivadas parciales de los componentes del campo} \end{cases}$$

—————> Anulación de las derivadas primeras $\partial/\partial u(\dots) = 0$; $\partial/\partial v(\dots) = 0$

Ecuaciones de Euler-Lagrange de minimización de C:

$$\begin{cases} 2(I_x u + I_y v + I_t)I_x + 2\lambda \left[\left(\frac{\partial}{\partial u} \frac{\partial u}{\partial x} \right) \frac{\partial u}{\partial x} + \left(\frac{\partial}{\partial u} \frac{\partial u}{\partial y} \right) \frac{\partial u}{\partial y} \right] = 0 \\ 2(I_x u + I_y v + I_t)I_y + 2\lambda \left[\left(\frac{\partial}{\partial v} \frac{\partial v}{\partial x} \right) \frac{\partial v}{\partial x} + \left(\frac{\partial}{\partial v} \frac{\partial v}{\partial y} \right) \frac{\partial v}{\partial y} \right] = 0 \end{cases}$$

$$\text{entonces: } \begin{cases} (I_x u + I_y v + I_t)I_x + \lambda \Delta u = 0 \\ (I_x u + I_y v + I_t)I_y + \lambda \Delta v = 0 \end{cases} \quad \text{con: } \begin{cases} \Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \\ \Delta v = \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \end{cases} \quad \text{laplacianos de } u \text{ y } v$$

Resolución de la EMA

2. Aproximación del laplaciano: $\nabla^2 f = f - \tilde{f}$ \tilde{f} = media de f dentro de cierta vecindad

pues:
$$\begin{cases} (v - \tilde{v})(I_x^2 + I_y^2 + \lambda) = -I_y(I_x \tilde{u} + I_y \tilde{v} + I_t) \\ (u - \tilde{u})(I_x^2 + I_y^2 + \lambda) = -I_x(I_x \tilde{u} + I_y \tilde{v} + I_t) \end{cases}$$

3. Resolución iterativa:

Método de Gauss-Seidel:
$$\begin{cases} u^k = \tilde{u}^{k-1} - I_x(I_x \tilde{u}^{k-1} + I_y \tilde{v}^{k-1} + I_t) \cdot (I_x^2 + I_y^2 + \lambda)^{-1} \\ v^k = \tilde{v}^{k-1} - I_y(I_x \tilde{u}^{k-1} + I_y \tilde{v}^{k-1} + I_t) \cdot (I_x^2 + I_y^2 + \lambda)^{-1} \end{cases}$$

Sea, tomando las notaciones originales:

ALGORITMO DE HORN & SCHUNCK

con:

$$\begin{cases} N = \frac{\partial I}{\partial x} \tilde{v}_x^{k-1} + \frac{\partial I}{\partial y} \tilde{v}_y^{k-1} + \frac{\partial I}{\partial t} \\ D = \lambda + \left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2 \end{cases}$$

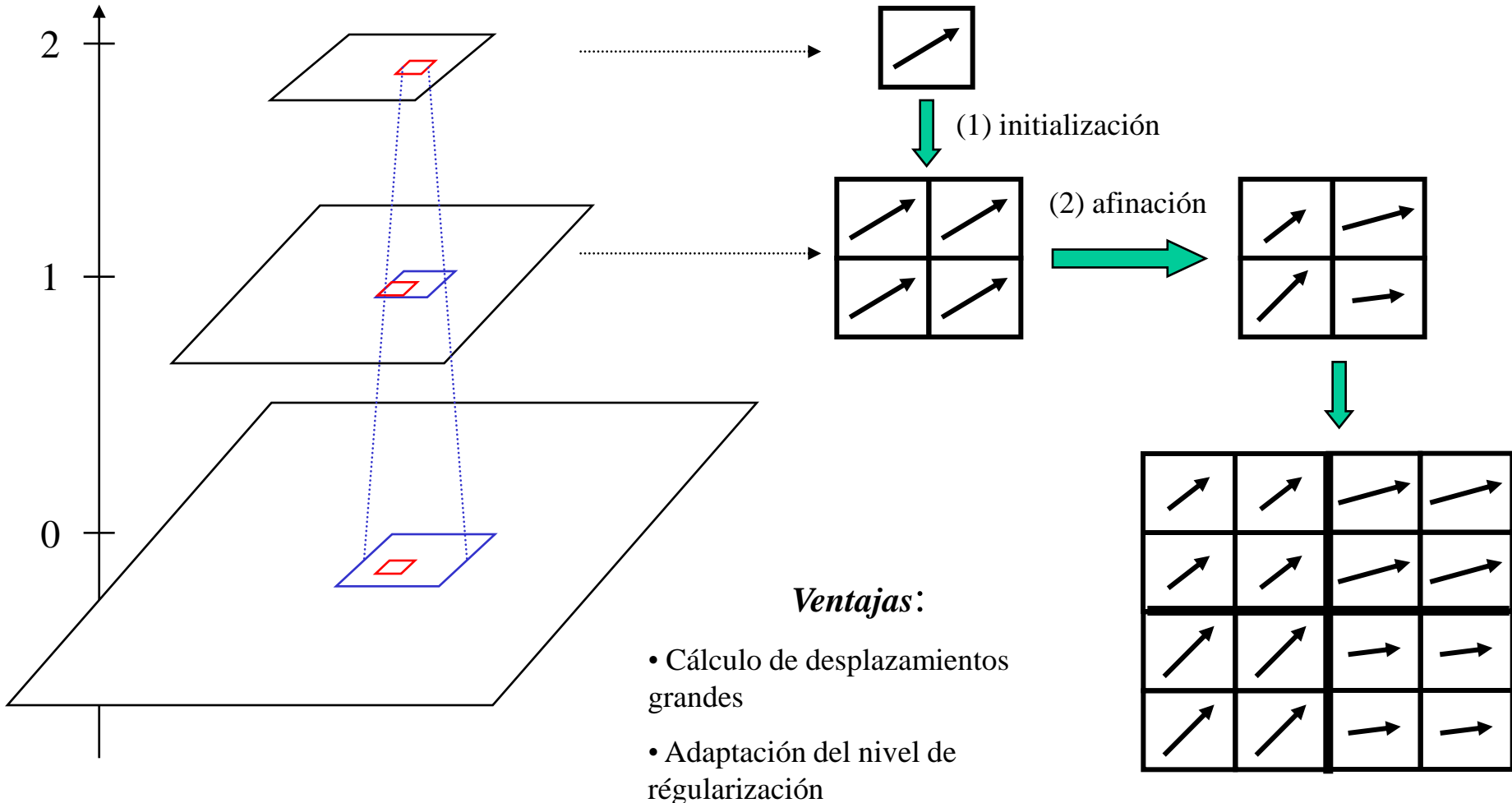
• Inicialización: $v_x^0 = 0$ $v_y^0 = 0$

• Repetir hasta convergencia:

$$\begin{cases} v_x^k = \tilde{v}_x^{k-1} - \frac{\partial I}{\partial x} \frac{N}{D} \\ v_y^k = \tilde{v}_y^{k-1} - \frac{\partial I}{\partial y} \frac{N}{D} \end{cases}$$

Estimación multi escala

Nivel



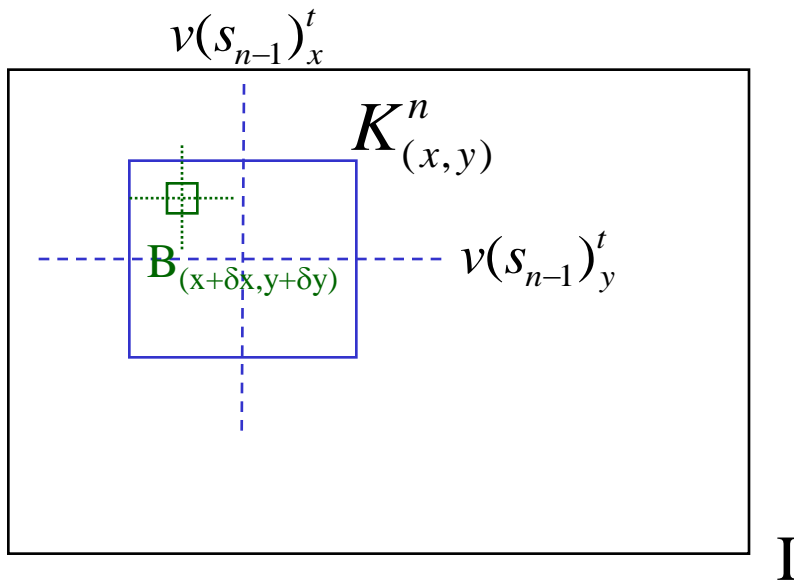
Ventajas:

- Cálculo de desplazamientos grandes
- Adaptación del nivel de regularización

Principio del calculo multi escala

ALGORITMO POR APAREAMIENTO

$$(v(s_n)_x, v(s_n)_y) = \arg \min_{(\delta x, \delta y) \in K^n} A^t_{(x,y)}(\delta x, \delta y)$$



$$\text{con: } K^n_{(x,y)} = \left[v(s_{n-1})_x \pm \Delta_1 \right] \times \left[v(s_{n-1})_y \pm \Delta_2 \right]$$

ALGORITMO DIFERENCIAL

- Inicialización: $v(s_0)_x^0 = 0$ $v(s_0)_y^0 = 0$
- Repetir hasta convergencia:

$$\left| \begin{array}{l} v(s_0)_x^k = \tilde{v}(s_0)_x^{k-1} - \frac{\partial I}{\partial x} \frac{N}{D} \\ v(s_0)_y^k = \tilde{v}(s_0)_y^{k-1} - \frac{\partial I}{\partial y} \frac{N}{D} \end{array} \right.$$

- Init. : $v(s_1)_x^0 = v(s_0)_x^\infty$ $v(s_1)_y^0 = v(s_0)_y^\infty$
- Repetir hasta convergencia:

$$\left| \begin{array}{l} v(s_1)_x^k = \tilde{v}(s_1)_x^{k-1} - \frac{\partial I}{\partial x} \frac{N}{D} \\ v(s_1)_y^k = \tilde{v}(s_1)_y^{k-1} - \frac{\partial I}{\partial y} \frac{N}{D} \end{array} \right.$$

.../...

Conclusión Capítulo 1

EXPLOTACION DEL FLUJO OPTICO

- Movimiento dominante
- Profundidad
- Tiempo antes colision

METODOS POR APAREAMIENTO

METODOS DIFERENCIALES

- Lucas et Kanade
- Horn et Schunck

METODOS MULTI-ESCALA

Bibliografía Capítulo 1

- **B.D. Lucas & T. Kanade 1981** « *An iterative image registration technique with an application to stereo vision* » *International Journal of Computer Vision and Artificial Intelligence* 674-679
- **B.K.P Horn & B. Schunck 1981** « *Determining Optical Flow* » *Artificial Intelligence* 23 185-203
- **D.H. Ballard & C.M Brown 1982** « *Computer Vision* » Prentice Hall (Ch. 3, Ch. 7)
- **J.M. Jolion & A. Rosenfeld 1994** « *A pyramid framework for early vision* » Kluwer Academic Publishers Dordrecht, NL
- **R. Jain, R. Kasturi & B. Schunck 1995** « *Machine Vision* » McGraw-Hill Inc. (Ch.14)

INDICE CAPITULO

- **Registracion: métodos frecuenciales**
- **seguimiento 1 : Harris y Invariantes de Hilbert**
- **seguimiento 2 : SIFT**

Introducción

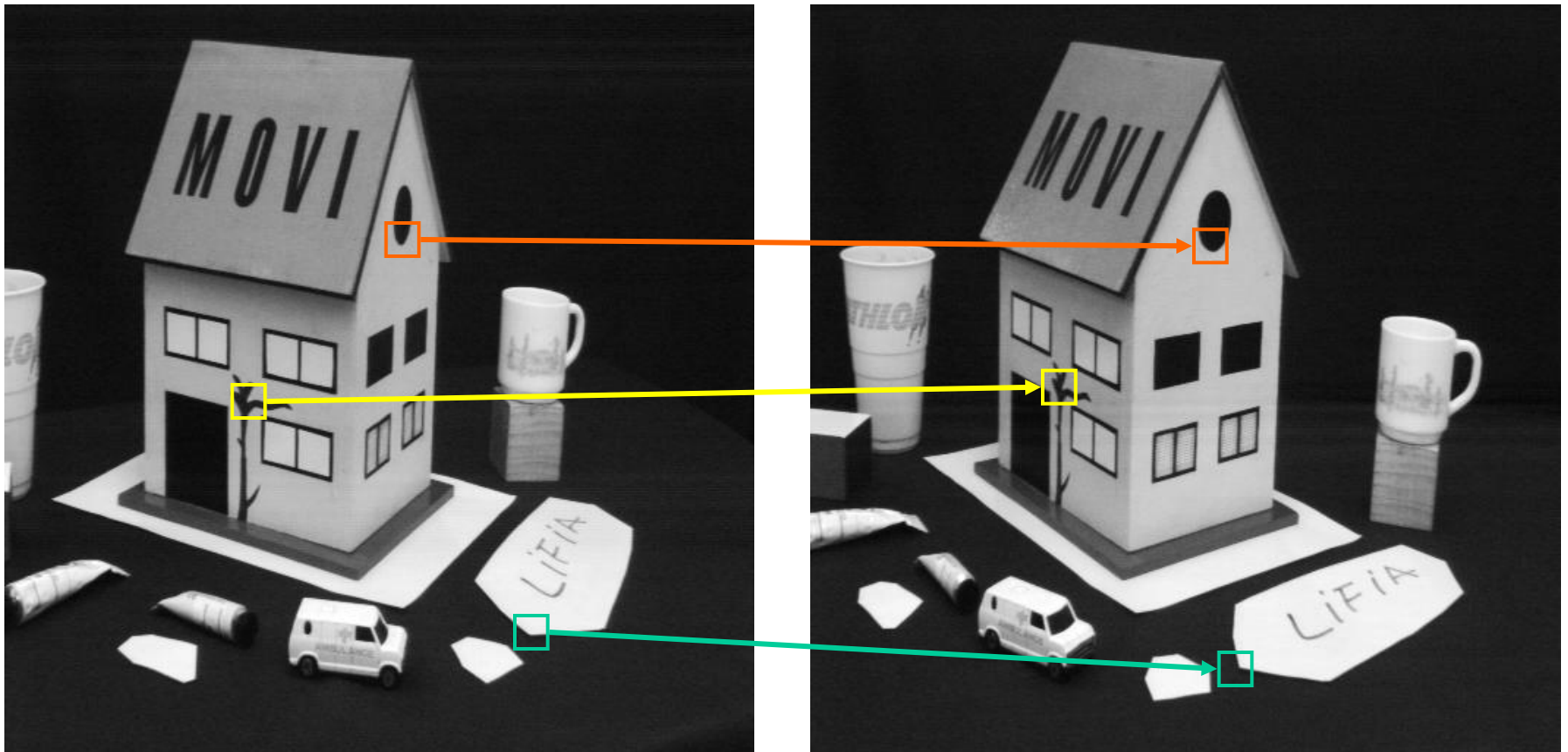


Cuando la amplitud del movimiento es grande, los métodos de estimación por apareamiento o diferenciales están condenadas al fracaso, por 2 motivos principales:

(1) El desplazamiento aparente puede ser superior a los períodos espaciales de las imágenes (aliasing temporal, véase el ejemplo a la izquierda)

Introducción

(2) *La escena está sujeta a distorsiones geométricas complejas que invalidan la hipótesis de traslación local (véase el ejemplo a continuación)*



Introducción

Por lo tanto:

- Es esencial disponer de descriptores que sean la mas *invariantes* posible a las *transformaciones geométricas*: rotación, homotecia, etc. Esto, inevitablemente, induce un cálculo multi-escala.
- Los descriptores tendrán carácter mucho *menos local*, e incluso global, pues ya no tendremos un campo denso como en los métodos de estimación, pero, en el mejor, un *campo disperso*.

Lo que vemos en este capítulo:

TRANSFORMACIÓN GLOBAL:

- Un solo parámetro para toda la imagen (traslación, rotación, homotecia): Métodos frecuenciales.

MOVIMIENTO CUALQUIER:

- ¿ Cuáles son los puntos utilizados para el cálculo ? Puntos de Harris, Maximos locales en el espacio de escala, ...
- ¿ Qué descriptores utilizar para el apareamiento ? Invariantes diferenciales, histogramas de orientación, ...

Aplicaciones: Reconocimiento de objetos, indexación de imágenes, ubicación de los robots,...

Técnicas frecuenciales

Las técnicas frecuenciales de estimacion de movimiento entre dos imágenes se basan en la equivalencia traslación / cambio de fase de la transformada de Fourier:

Recuerde: la expresión de una imagen en el dominio frecuencial consiste en descomponer la función de dos dimensiones en sumas de funciones senos complejas:

$$I(x, y) = \frac{1}{wh} \sum_{u=0}^{w-1} \sum_{v=0}^{h-1} F(u, v) e^{2i\pi(ux+vy)/wh}$$

Transformada de Fourier discreta inversa



Los coeficientes de los diferentes senos se calculan por la transformada de Fourier:

$$F(u, v) = \sum_{x=0}^{w-1} \sum_{y=0}^{h-1} I(x, y) e^{-2i\pi(ux+vy)/wh}$$

Transformada de Fourier discreta directa

Notacion (modulo, fase) : $F(u, v) = \|F(u, v)\| e^{i\varphi_F(u, v)}$

Propiedad de traslacion/cambio de fase: si F es la Transformada de Fourier (TF) de I :

Entonces la TF de I traslada de $(-\delta x, -\delta y)$, es G , con:

$$\begin{array}{ccc} I(x, y) & \xrightarrow{\text{TF}} & F(u, v) \\ I(x + \delta x, y + \delta y) & \xrightarrow{\text{TF}} & G(u, v) = F(u, v) e^{2i\pi(u\delta x + v\delta y)/wh} \end{array}$$

pues: $\|G(u, v)\| = \|F(u, v)\|$ $\varphi_G(u, v) = \varphi_F(u, v) + 2\pi(u\delta x + v\delta y)/wh$

La diferencia de fase entre F y G es: $\Delta\phi(u, v) = 2\pi(u\delta x + v\delta y)/wh$

Por lo tanto, es suficiente en teoría considerar esta diferencia de fase para 2 parejas (u, v) , para calcular (δ_x, δ_y) , pero esta técnica es sensible al ruido y a los cambios de iluminación que inducen cambios en las frecuencias bajas.

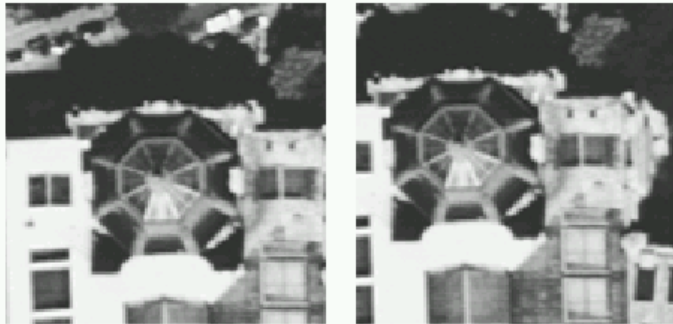
Se utiliza más bien la técnica de correlación de fase.

Correlación de fase

La técnica de correlación de fase explota una consecuencia directa de la propiedad de la traslación / cambio de fase. Si F es la TF de I y G la TF de I trasladada de $(-\delta x, -\delta y)$, entonces la diferencia entre F y G es igual a su espectro de potencia cruzado normalizado (PSNA), es decir, :

$$\frac{F^*(u, v)G(u, v)}{\|F^*(u, v)G(u, v)\|} = e^{2i\pi(u\delta x + v\delta y)/wh}$$

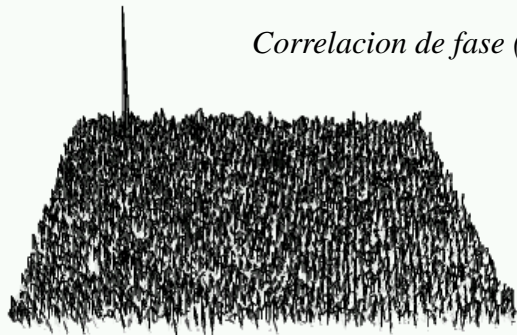
La TF inversa del PSNA es igual entonces a la función de Dirac del vector de traslación: $\delta_{(\delta x, \delta y)}(x, y)$



(a)

(b)

Correlacion de fase (INRIA)



(c)

La técnica de correlación de fase consiste entonces en:

1. Calcular las TF de $I(x, y, t)$ y $I(x, y, t+1)$, sea F_1 et F_2
2. Calcular χ el PSNA de F_1 y F_2
3. Calcular D la TF inversa de χ
4. Buscar al máximo de D

Ventajas y inconvenientes

- + Robusto por qué todas las frecuencias contribuyen al calculo
- + Relativamente rapido gracia al calculo de la FFT
- En practica limitado a un desplazamiento global sobre la imagen entera

Invariantes de Fourier-Mellin

El uso de la transformada de *Fourier-Mellin* permite calcular los parámetros de una similitud (*homotecia y rotación*) como un *vector de traslación* de una manera similar al caso anterior, a través de una representación log-polar del espacio de frecuencias espaciales $(u, v) \rightarrow (\theta, \log \rho)$:

Sea g la imagen transformada de f por una rotación de ángulo α , una homotecia de razón σ , y una traslación de vector (x_0, y_0) :

$$g(x, y) = f(\sigma(\cos\alpha x + \sin\alpha y) - x_0, \sigma(-\sin\alpha x + \cos\alpha y) - y_0)$$

Los módulos de las transformadas de Fourier de f y g son vinculados por la relación siguiente:

$$\|G(u, v)\| = \frac{1}{\sigma^2} \|F(\frac{1}{\sigma}(u \cos \alpha + v \sin \alpha), \frac{1}{\sigma}(-u \sin \alpha + v \cos \alpha))\|$$

Pues el módulo: $\left\{ \begin{array}{l} \bullet \text{ no depende de la traslación } (x_0, y_0). \\ \bullet \text{ sufre una rotación de ángulo } \alpha. \\ \bullet \text{ sufre un cambio de escala de factor } 1/\sigma. \end{array} \right.$

Pasando las frecuencias en coordenadas polares:

$$F_p(\theta, \rho) = \|F(\rho \cos \theta, \rho \sin \theta)\|; 0 \leq \theta \leq 2\pi, 0 \leq \rho < \infty$$

$$G_p(\theta, \rho) = \|G(\rho \cos \theta, \rho \sin \theta)\|; 0 \leq \theta \leq 2\pi, 0 \leq \rho < \infty$$

Por fin, pasando la coordenada radial al logaritmo:

$$r = \log \rho$$

$$s = \log \sigma$$

$$F_{lp}(\theta, r) = F_p(\theta, \rho)$$

$$G_{lp}(\theta, r) = G_p(\theta, \rho)$$

tenemos:

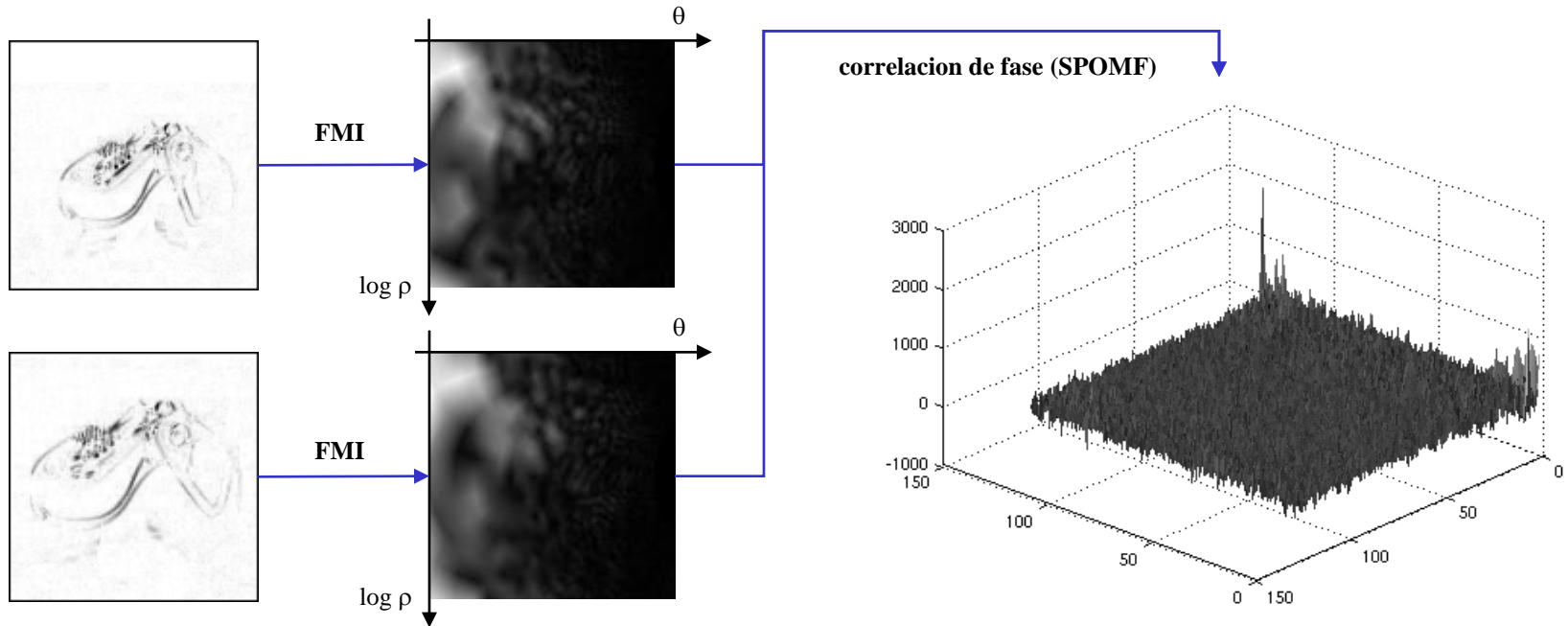
$$G_p(\theta, \rho) = \frac{1}{\sigma^2} F_p\left(\theta - \alpha, \frac{\rho}{\sigma}\right)$$

tenemos:

$$G_{lp}(\theta, r) = \frac{1}{\sigma^2} F_{lp}(\theta - \alpha, r - s)$$

Entonces la similitud en el dominio espacial se traduce por una *traslación* en el espacio de *frecuencias log-polares*.

Invariantes de Fourier-Mellín



Un ejemplo de uso de la transformada de Fourier-Mellin: cálculo de la posición de la cabeza de robots Aibo en la imagen por correlación de fase de los invariantes de Fourier-Mellin. (FMI-SPOMF : Fourier-Mellin Invariant Symmetric Phase Only Matched Filtering) : *J.C. Baillie et M. Nottale* 2004.



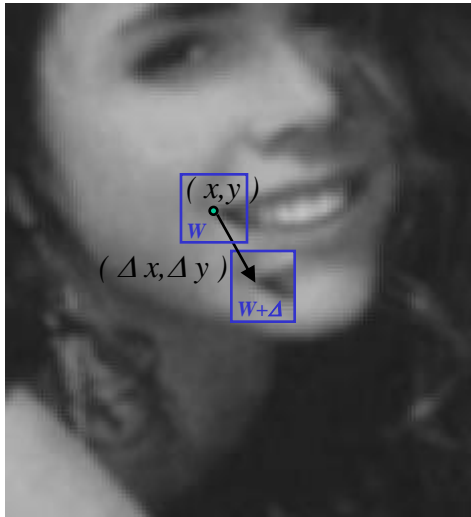
La información de fase en la imagen original se pierde en el FMI. El FMI-SPOMF equivale a buscar la mejor (rotación, homotecia) que pone en correspondancia los 2 espectros de amplitud. *Por lo tanto, no refleja los parámetros de traslación entre las dos imágenes, y además, la información llevada por la fase ya no existe.*

Para completar esta transformación, podemos aplicar una correlación de fase clásica sobre la pareja de imágenes iniciales, después de aplicarse a una de las imágenes la transformación (rotación, dilatación) proporcionado por el FMI-SPOMF.

Por último, observe que, al igual de la correlación de fase, este método se utiliza en la práctica para estimar las *transformaciones globales*, porque utiliza la contribución de todo el espectro (o por lo menos de una gran parte del mismo), lo que supone una importante extensión espacial de los píxeles usados para estimar cada transformación.

Deteccion de puntos de interés

Los puntos de interés (o puntos angulosos, puntos salientes,...), son puntos que « contienen gran cantidad de información » en relación con la imagen. Son los puntos al vecindad de los cuales la imagen *varía considerablemente en varias direcciones*.



Una medida de las variaciones locales de la imagen I en el punto (x, y) asociada a un desplazamiento $(\Delta x, \Delta y)$ es la *funcion de autocorrelacion*:

$$\chi(x, y) = \sum_{(x_k, y_k) \in W} (I(x_k, y_k) - I(x_k + \Delta x, y_k + \Delta y))^2$$

Donde W es una ventana centrada en el punto (x, y) .

Pues, usando una aproximacion de la primera orden:

$$I(x_k + \Delta x, y_k + \Delta y) \approx I(x_k, y_k) + \left(\frac{\partial I}{\partial x}(x_k, y_k) \quad \frac{\partial I}{\partial y}(x_k, y_k) \right) \cdot \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$$

Entonces:

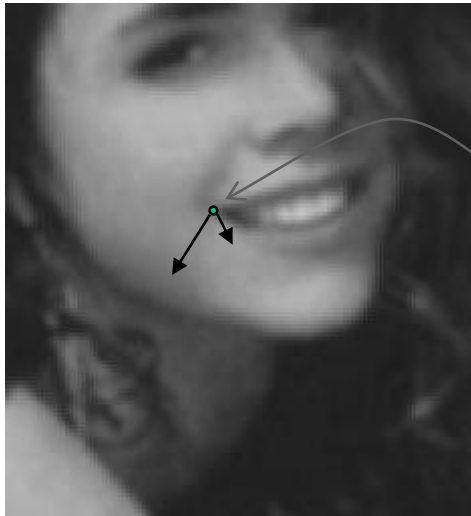
$$\chi(x, y) = \sum_{(x_k, y_k) \in W} \left(\left(\frac{\partial I}{\partial x}(x_k, y_k) \quad \frac{\partial I}{\partial y}(x_k, y_k) \right) \cdot \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \right)^2 = (\Delta x \quad \Delta y) \underbrace{\begin{pmatrix} \sum_{(x_k, y_k) \in W} \left(\frac{\partial I}{\partial x}(x_k, y_k) \right)^2 & \sum_{(x_k, y_k) \in W} \frac{\partial I}{\partial x}(x_k, y_k) \cdot \frac{\partial I}{\partial y}(x_k, y_k) \\ \sum_{(x_k, y_k) \in W} \frac{\partial I}{\partial x}(x_k, y_k) \cdot \frac{\partial I}{\partial y}(x_k, y_k) & \sum_{(x_k, y_k) \in W} \left(\frac{\partial I}{\partial y}(x_k, y_k) \right)^2 \end{pmatrix}}_{\Xi(x, y)} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$$

Matriz de autocorrelacion de la imagen I en (x, y)

Deteccion de puntos de interés

$$\Xi(x, y) = \begin{pmatrix} \sum_{(x_k, y_k) \in W} \left(\frac{\partial I}{\partial x}(x_k, y_k) \right)^2 & \sum_{(x_k, y_k) \in W} \frac{\partial I}{\partial x}(x_k, y_k) \cdot \frac{\partial I}{\partial y}(x_k, y_k) \\ \sum_{(x_k, y_k) \in W} \frac{\partial I}{\partial x}(x_k, y_k) \cdot \frac{\partial I}{\partial y}(x_k, y_k) & \sum_{(x_k, y_k) \in W} \left(\frac{\partial I}{\partial y}(x_k, y_k) \right)^2 \end{pmatrix}$$

La matriz de autocorrelacion Ξ representa la variacion local de la imagen I en (x, y) . (x, y) sera considerado punto de interés de I si para todos los desplazamientos $(\Delta x, \Delta y)$, la cantidad $(\Delta x, \Delta y) \cdot \Xi(x, y) \cdot (\Delta x, \Delta y)^t$ es grande.



Los puntos de interés son los puntos (x, y) para los cuales la matriz de autocorrelacion $\Xi(x, y)$ tiene *dos autovalores grandes*.

Esto corresponde a los puntos para los cuales existe localmente una base de autovectores de Ξ describiendo variaciones locales importantes de la imagen.

El *detector de Harris* calcula una *funcion de interés* $\Theta(x, y)$:

$$\Theta(x, y) = \det \Xi - \alpha \text{traza} \Xi$$

El primer término corresponde al producto de las autovalores, el segundo término penaliza los puntos de contorno con una sola fuerte autovalor grande.

Los puntos de interés corresponden a los maximos locales de la funcion Θ cuya valor supera un cierto umbral (tipicamente el 1% de la valor maxima de Θ).

Utilización de los invariantes diferenciales

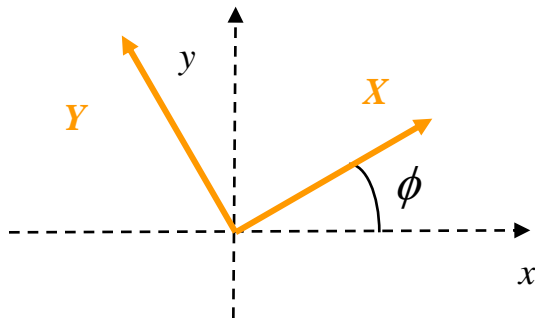
Objetivo : representar los puntos de interes por *indices* que sean *invariantes* por *rotacion* y *cambio de escala*.

El principio usado aqui es basado sobre el calculo de las derivadas multi escala:

El «local jet» de I : $L_{ij}^\sigma = I * G_{ij}^\sigma$ con: $G_{ij}^\sigma = \frac{\partial^{i+j}}{\partial x^i \partial y^j} G^\sigma$ y: $G^\sigma(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$

Se denota: $\{L_{ij}^\sigma; 0 \leq i + j \leq 3\} = \{L, L_x, L_y, L_{xx}, L_{xy}, L_{yy}, L_{xxx}, L_{xxy}, L_{xyy}, L_{yyy}\}$ (derivadas hasta la orden 3)

La idea es *combinar estas derivadas* para obtener cantidades *invariantes por rotacion*:



Por ejemplo, el laplaciano $I_{xx} + I_{yy}$ es invariante por rotacion:

$$\begin{cases} x = X \cos \phi + Y \sin \phi \\ y = X \sin \phi - Y \cos \phi \end{cases} \quad \begin{cases} X = x \cos \phi + y \sin \phi \\ Y = -x \sin \phi + y \cos \phi \end{cases}$$

$$\begin{cases} I_X = I_x \cos \phi + I_y \sin \phi \\ I_Y = I_x \sin \phi - I_y \cos \phi \end{cases} \quad \text{y} : \begin{cases} I_{XX} = I_{xx} \cos^2 \phi + 2I_{xy} \cos \phi \sin \phi + I_{yy} \sin^2 \phi \\ I_{YY} = I_{xx} \sin^2 \phi - 2I_{xy} \cos \phi \sin \phi + I_{yy} \cos^2 \phi \end{cases}$$

entonces: $I_{XX} + I_{YY} = I_{xx} + I_{yy}$

Utilización de los invariantes diferenciales

Así se puede construir una familia de cantidades invariantes por rotación: los *invariantes diferenciales de Hilbert*.

$$\Psi = \left(\begin{array}{c} L \\ L_i L_i \\ L_i L_{ij} L_j \\ L_{ii} \\ L_{ij} L_{ij} \\ \varepsilon_{ij} (L_{jkl} L_i L_k L_l - L_{jkk} L_i L_l L_l) \\ L_{ijj} L_j L_k L_k - L_{ijk} L_i L_j L_k \\ - \varepsilon_{ij} L_{jkl} L_i L_k L_l \\ L_{ijk} L_i L_j L_k \end{array} \right)$$

$$\text{Con: } \begin{array}{l} \varepsilon_{xx} = \varepsilon_{yy} = 0 \\ \varepsilon_{xy} = -\varepsilon_{yx} = 1 \end{array}$$

(notaciones de Einstein: sumación sobre los índices), por ejemplo:

$$\begin{aligned} \Psi_2 &= L_i L_{ij} L_j = L_x L_{xx} L_x + 2L_x L_{xy} L_y + L_y L_{yy} L_y \\ \Psi_7 &= -\varepsilon_{ij} L_{jkl} L_i L_k L_l = L_{xxy} (-L_x L_x L_x + 2L_x L_y L_y) \\ &+ L_{xyy} (-2L_x L_x L_y + L_y L_y L_y) - L_{yyy} L_x L_y L_y + L_{xxx} L_x L_x L_y \end{aligned}$$

NB: ¡ invariancia por rotación del núcleo gaussiano !

Así, los vectores Ψ son calculados por todos los puntos de interés a diferentes escalas, y apareados utilizando una distancia (e.g. distancia euclidiana).

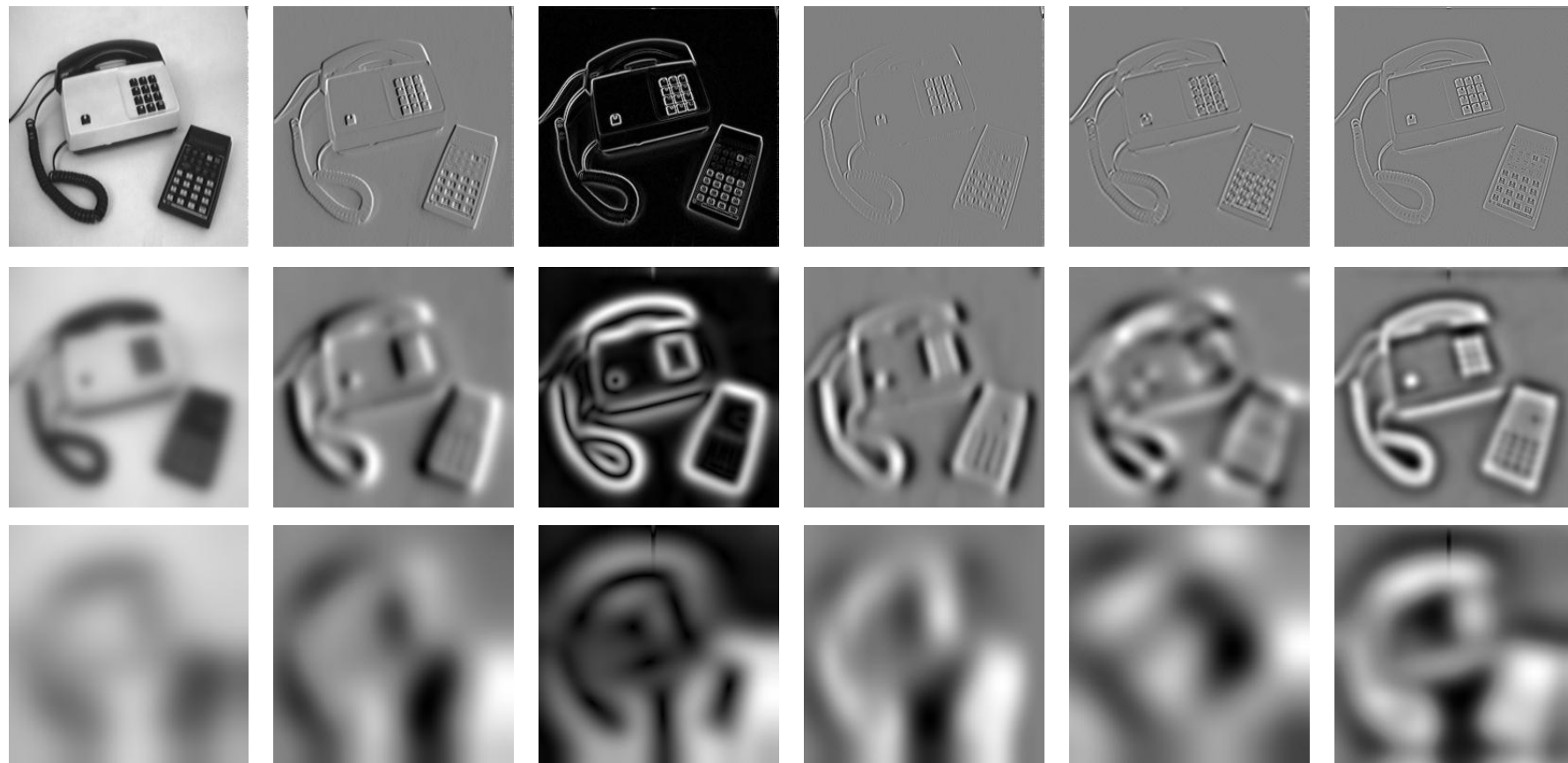
Espacio de escalas gaussiano y derivadas

$$g_\sigma(x) = \frac{1}{\sqrt{\pi}\sigma} \exp\left(-\frac{x^2}{\sigma^2}\right)$$

$$\frac{\partial g_\sigma}{\partial x}$$

$$\frac{\partial^2 g_\sigma}{\partial x^2}$$

$$\partial^n (I * g_\sigma) = I * \partial^n g_\sigma$$



I

$$\frac{\partial I}{\partial x}$$

$$\|\nabla I\| = \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2}$$

$$\frac{\partial^2 I}{\partial x^2}$$

$$\frac{\partial^2 I}{\partial x \partial y}$$

$$\Delta I = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$$

escala

Aplicación del detector de Harris

1. Se calcula las derivadas primeras a partir de convoluciones con derivadas de gaussiana (desviación estándar σ_1)
2. Se calcula los componentes de la matriz de autocorrelación Ξ calculando una media local de las derivadas bajo la forma de una gaussiana (desviación estándar σ_2 , típicamente $\sigma_2 = 2 \sigma_1$)
3. Se calcula la función de interés: $\Theta = \det(\Xi) - \alpha \text{traza}(\Xi)$ (típicamente $\alpha = 0,06$).
4. Se calcula los máximos locales de Θ superiores a un cierto umbral (típicamente el 1% de Θ_{\max}).

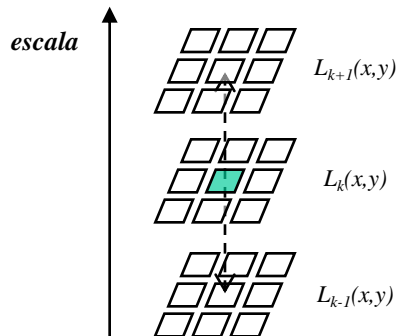
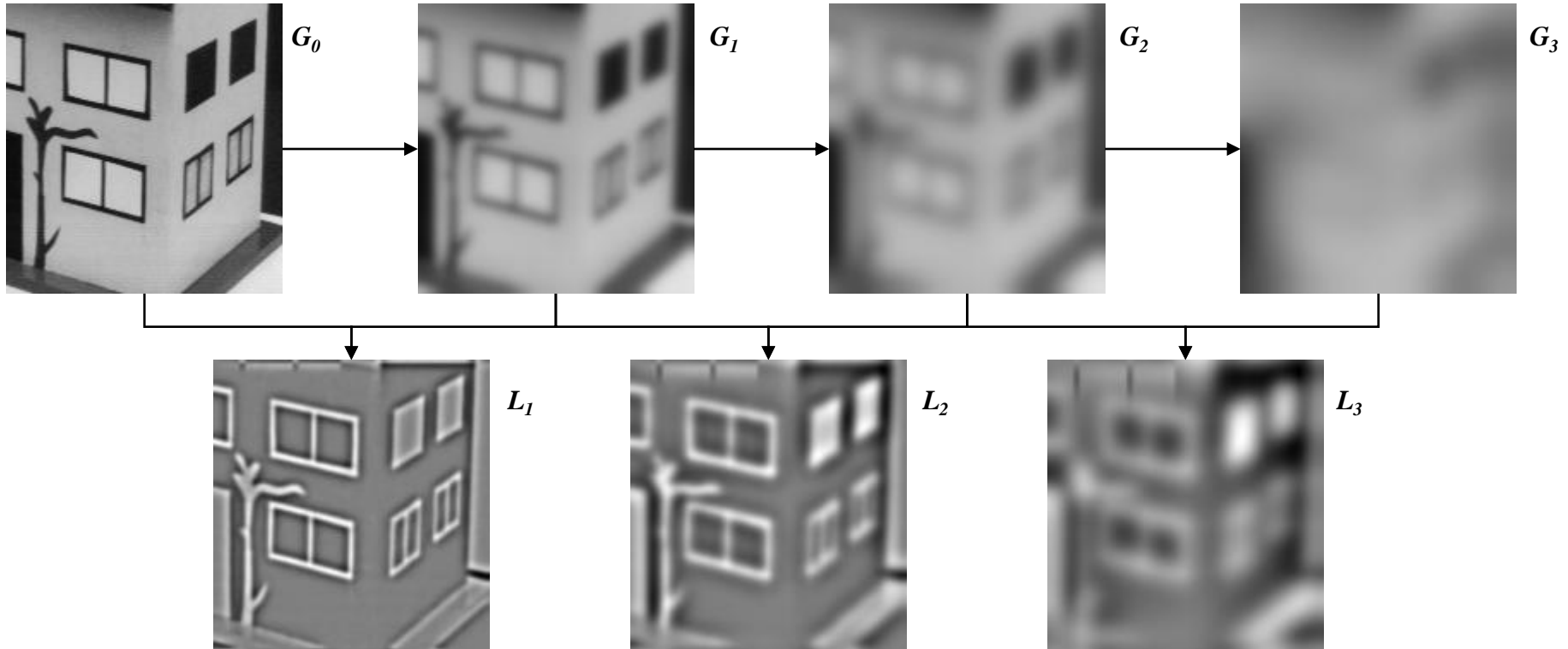


Detector de Harris multi escala



Puntos de Harris obtenidos calculando las derivadas primeras por convolucion con una derivada de gaussiana de desviacion estandar σ .

Detector SIFT: extrema en el espacio de escala



La función $G_k(x,y) = G(x,y,k\sigma)$ es la imagen suavizada por una gaussiana de desviación estandar $k\sigma$. Las funciones $L_k(x,y)$ corresponden a la diferencia (normalizada aquí) entre 2 gaussianas consecutivas.

La función $L_k(x,y)$ es una representación laplaciana de la imagen, es decir una descomposición frecuencial localizada: contribución de las estructuras de escala (tamaño) $k\sigma$ en el punto (x,y) .

Los puntos seleccionados por SIFT son los máximos y mínimos locales de la función $L_k(x,y)$, a la vez en la escala corriente y en las escalas adyacentes (vease a la izquierda).

Puntos de interés SIFT

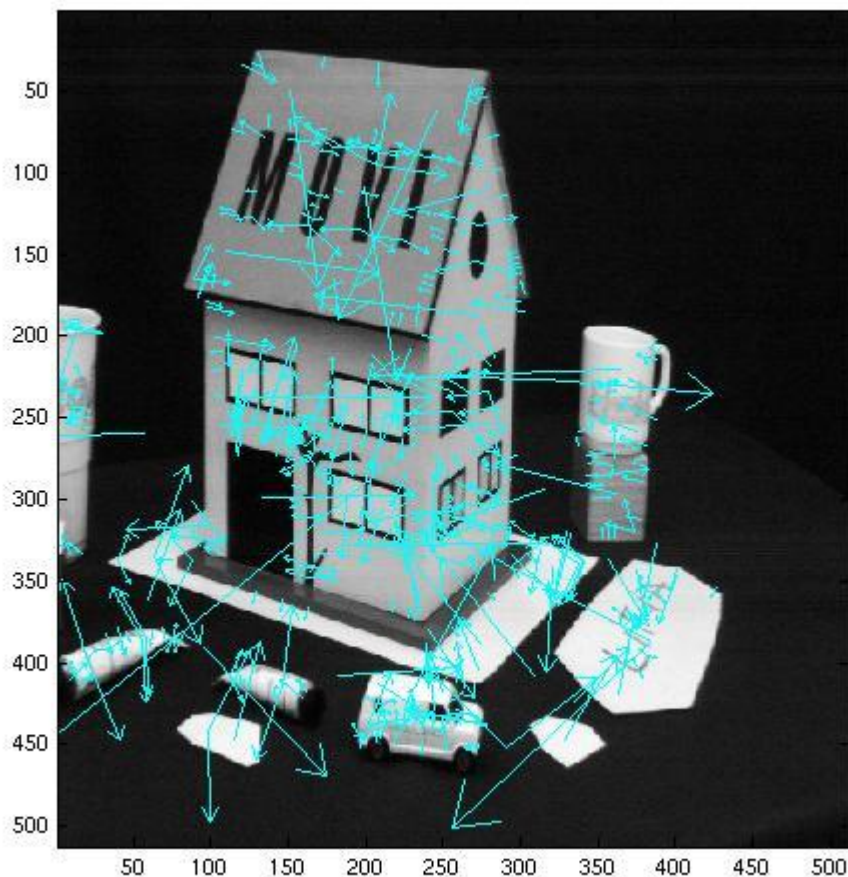


Imagen 1: 589 puntos detectados.

Por cada extremo del espacio de escala de las diferencias de gaussianas (punto de interés SIFT), se calcula la dirección asociada de la manera siguiente:

$$\theta(x, y) = \arctan\left(\frac{G_y^\sigma(x, y)}{G_x^\sigma(x, y)}\right)$$

$$\text{con } G_x^\sigma(x, y) = \frac{\partial}{\partial x} G(x, y, \sigma) = I(x, y) * \frac{\partial}{\partial x} g_\sigma(x, y)$$

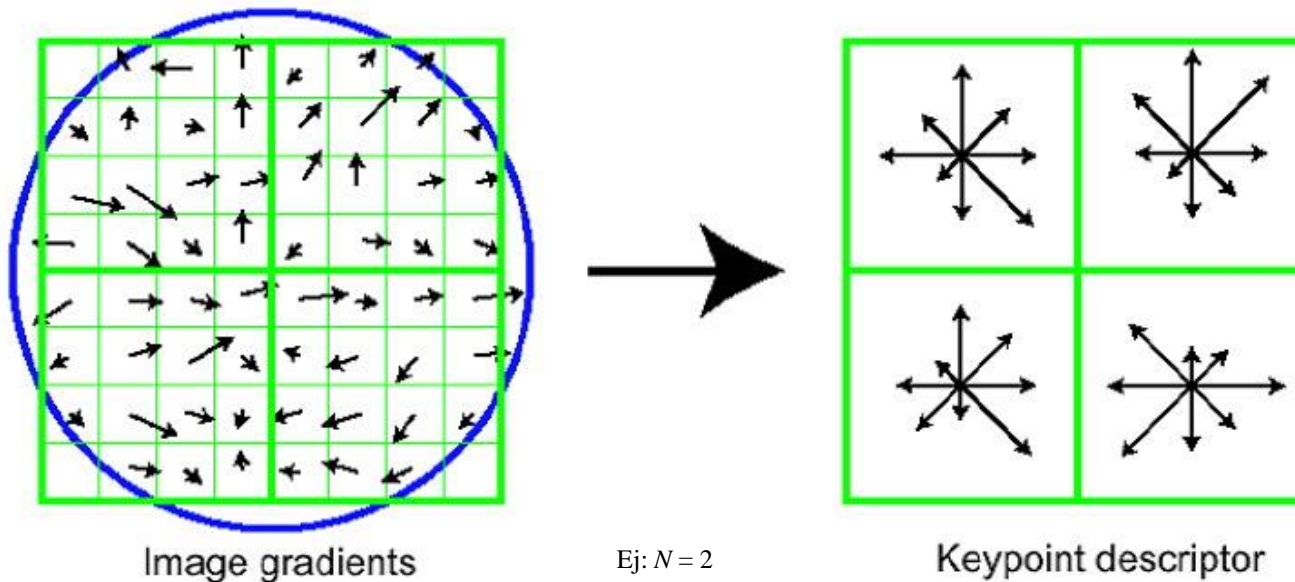
(donde σ es la escala seleccionada)

Al la izquierda, puntos de interés SIFT: la dirección de la flecha representa la dirección θ et su largura la escala σ asociada.

Descriptor SIFT: histograma de orientacion

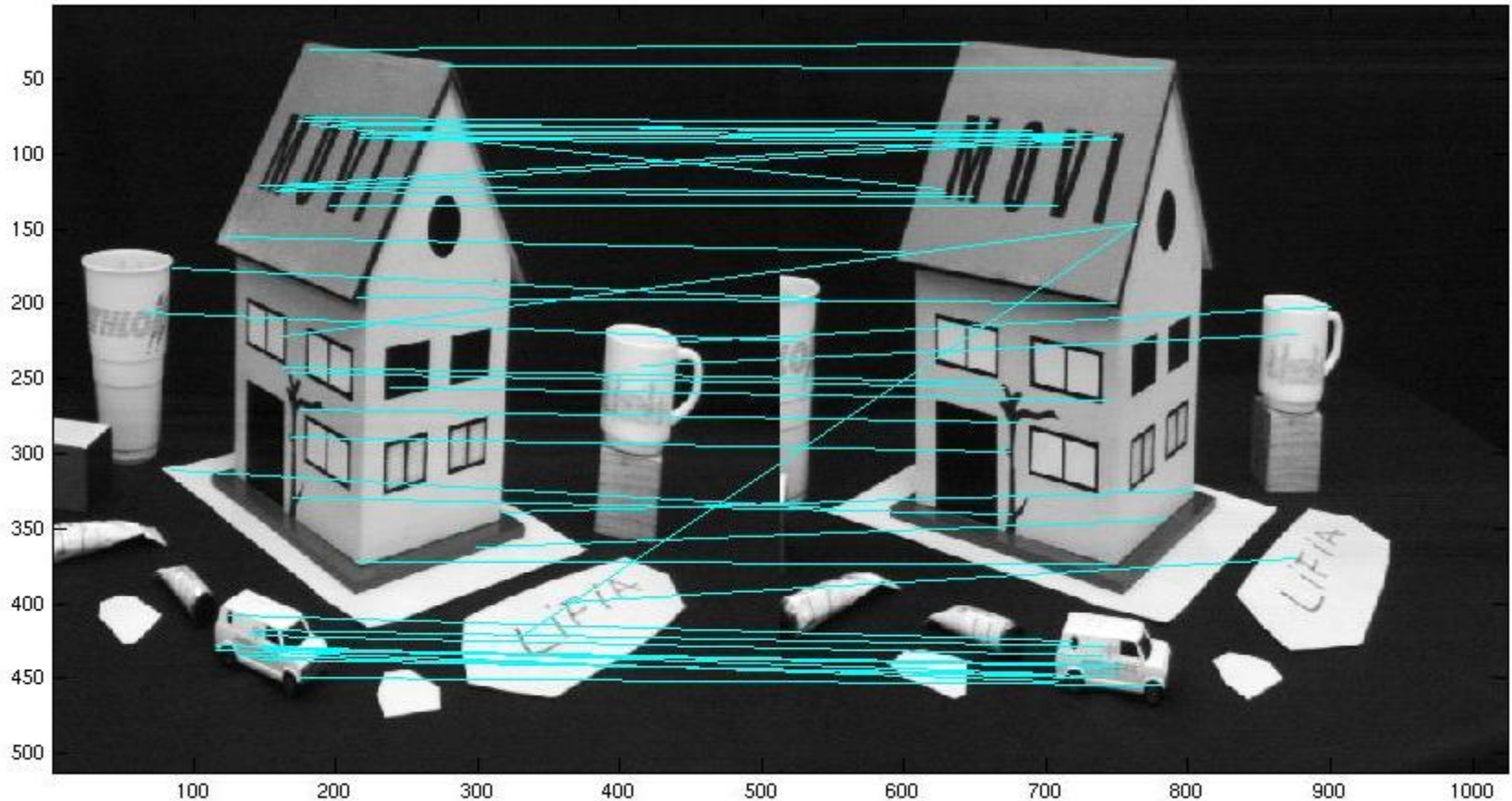
Los descriptores asociados a los puntos SIFT son histogramas de las orientaciones locales alrededor del punto.

- Se divide el espacio alrededor de cada punto de interés (x,y) en N^2 cuadrados 4×4 .
- Se calcula el gradiente $(G_x(a,b,\sigma), G_y(a,b,\sigma))$ para los $4 \times 4 \times N^2$ puntos (a,b) .
- Para cada cuadrado 4×4 , se calcula un histograma de las orientaciones cuantificadas en 8 direcciones, ponderando los valores por: (1) el modulo del gradiente (2) la inversa de la distancia al punto de interés (x,y) .
- Para lograr invariancia en rotacion: la orientacion local del punto de interés $\theta(x,y)$ esta usada como *origen* (orientacion nula) de los histogramas.



Pues los descriptores formados son vectores de tamaño $8 \times N^2$, que seran apareados usando una distancia (e.g. distancia euclidiana)

Apareamiento por SIFT



Resultado de apareamiento por SIFT entre la imagen (2) a la izquierda, 510 puntos detectados, y la imagen (1) a la derecha, 589 puntos detectados. 51 puntos fueron apareados, lo que corresponde a una distancia euclidiana entre los descriptores por debajo de cierto umbral.

Conclusión Capítulo 2

- ❑ REGISTRACIÓN: Movimiento global
 - ❑ Correlacion de fase: *traslación*
 - ❑ Invariantes de Fourier-Mellin: *rotación / escalamiento*
- ❑ SEGUIMIENTO
 - ❑ Puntos de interés
 - ❑ Harris: *puntos angulosos*
 - ❑ SIFT: *extremos en el espacio de escala*
 - ❑ Descriptores
 - ❑ Invariantes diferenciales
 - ❑ Histogramas de orientación

Bibliografía del Capítulo 2

- **H. Foroosh, J. Zerubia & M. Berthod 2002** « *Extension of phase correlation to subpixel registration* » *IEEE Transactions on Image Processing* 11(3) pp 188-200
- **Q. Chen, M. Defrise & F. Deconinck 1994** « *Symmetric Phase-Only Matched Filtering of Fourier-Mellin Transforms for Image Registration and Recognition* » *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16(12) pp 1156-1168
- **C. Harris & M. Stephens 1988** « *A combined corner and edge detector* » *Alvey Vision Conference* pp 147-151
- **C. Schmid & R. Mohr 1997** « *Local grayvalue invariants for image retrieval* » *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(5) pp 530-534
- **D.G. Lowe 2004** « *Distinctive Image Features from Scale-Invariant Keypoints* » *International Journal of Computer Vision* 60-2 pp 91-110